SIAM J. MATH. ANAL.
Vol. xx, No. x, pp. x–x

THE JUMP SET UNDER GEOMETRIC REGULARISATION. PART 1: BASIC TECHNIQUE AND FIRST-ORDER DENOISING

TUOMO VALKONEN*

Abstract. Let $u \in \mathrm{BV}(\Omega)$ solve the total variation denoising problem with L^2-squared fidelity and data f. Caselles et al. [Multiscale Model. Simul. 6 (2008), 879–894] have shown the containment $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$ of the jump set J_u of u in that of f. Their proof unfortunately depends heavily on the co-area formula, as do many results in this area, and as such is not directly extensible to higher-order, curvature-based, and other advanced geometric regularisers, such as total generalised variation (TGV) and Euler's elastica. These have received increased attention in recent times due to their better practical regularisation properties compared to conventional total variation or wavelets. We prove analogous jump set containment properties for a general class of regularisers. We do this with novel Lipschitz transformation techniques, and do not require the co-area formula. In the present Part 1 we demonstrate the general technique on first-order regularisers, while in Part 2 we will extend it to higher-order regularisers. In particular, we concentrate in this part on TV and, as a novelty, Huber-regularised TV. We also demonstrate that the technique would apply to non-convex TV models as well as the Perona-Malik anisotropic diffusion, if these approaches were well-posed to begin with.

AMS subject classifications. 26B30, 49Q20, 65J20.

Key words. total variation, jump set, regularisation, Lipschitz, Huber, Perona-Malik.

1. Introduction.

We study the structure of the approximate jump set J_u of solutions $u \in \mathrm{BV}(\Omega)$ to regularisation problems

$$\min_{u \in L^1(\Omega)} \int_\Omega \phi(f(x) - u(x))\,dx + R(u). \tag{P}$$

We recall that J_u is the $\mathcal{H}^{m-1}$-rectifiable set on which u has two distinct one-sided Lebesgue limits. We consider domains $\Omega \subset \mathbb{R}^m$ for $m \geq 2$, and assume that $\phi : [0,\infty) \to [0,\infty]$ is a convex, lower semi-continuous, p-increasing fidelity function, and R a regularisation functional, which generalises total variation (TV) in a suitable sense. The technical conditions that we set on R are to ensure that solutions satisfy $u \in \mathrm{BV}(\Omega)$ and that R behaves almost like TV under small Lipschitz transformations. We state these conditions in detail in §3. Briefly, we require the BV-coercivity condition

$$\|Du\|_{2,\mathcal{M}(\Omega;\mathbb{R}^m)} \leq C\big(1 + \|u\|_{L^1(\Omega)} + R(u)\big), \quad (u \in L^1(\Omega)),$$

and a *double-Lipschitz comparability condition* of the form

$$R(\overline{\gamma}_\# u) + R(\underline{\gamma}_\# u) - 2R(u) \leq C T_{\overline{\gamma},\underline{\gamma}} |Du|(\operatorname{cl} U). \tag{1.1}$$

Here $\overline{\gamma}$ and $\underline{\gamma}$ are Lipschitz transformations on an open set U. The suitably defined distance $T_{\overline{\gamma},\underline{\gamma}}$ between the transformations turns out to be $O(\rho^2)$ for specially constructed shift transformations, dependent on a parameter ρ. As we will see in §4, this class of regularisation functionals includes, in particular, total variation (TV) and Huber-regularised total variation. In Part 2 [41] of this pair of papers, we will look

* Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK. The most significant part of this work was done while the author was a Prometeo Fellow at the Research Center on Mathematical Modeling (Modemat), Escuela Politécnica Nacional de Quito, Ecuador. E-mail: tuomo.valkonen@iki.fi.

at the extension of this condition and technique to cover higher-order regularisers, in particular total generalised variation (TGV) [6], and infimal convolution TV (ICTV) [11]. The analysis of these functionals, in particular that of TGV, is significantly more involved than that of first-order regularisers, and enough to fill an additional manuscript or three.

Assuming further that both the solution u and the given data f are in $\mathrm{BV}(\Omega) \cap L^\infty_{\mathrm{loc}}(\Omega)$, we show for p-increasing ϕ for $1 < p < \infty$, including L^p fidelities $\phi(t) = |t|^p$, that the jump set J_u of u is contained, modulo a $\mathcal{H}^{m-1}$ null set, in the jump set J_f of f. That is, $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$. The boundedness condition of course holds for TV and Huber-TV by standard barrier arguments, but has to be shown or imposed separately in the general case. For $\phi(x) = x$, i.e., the L^1 fidelity, the same conclusion does not necessarily hold, as is known in the case of total variation regularisation [18]. Under assumptions of approximate piecewise constancy, we however show that $J_u \setminus J_f$ has a $C^{2,\gamma}$ structure with curvature $1/\alpha$, for α the (asymptotic) regularisation parameter. We state all of these results in detail in §3. The proofs are split between multiple sections. We construct the specific Lipschitz transformations in §5. The main part of the general proof, studying the effect of the transformations on the fidelity, can be found in §6 for $p > 1$, and in §7 for $p = 1$.

The class of problems (P) is of importance, in particular, for image denoising. From an application point of view, it is desirable to know the structure of J_u in order to show that the regularisation method is reliable – that it does not introduce undesirable artefacts, new edges, and correctly restores edges where they are present in the original data f. Higher-order geometric regularisation functionals, such a total generalised variation (TGV) [6] and infimal convolution TV [11], are as a matter of fact motivated by other artefacts introduced by TV regularisation: the stair-casing effect. Total variation exhibits flat areas with sharp transitions, which higher-order regularisation tends to avoid. We also note that while the L^2 fidelity $\phi(t) = t^2$, is often easier from the computational point of view, the L^1 fidelity can better deal with outliers. The L^2 fidelity models Gaussian noise, while the L^1 fidelity closer models impulse noise, and should therefore be preferred from the point of view of robust statistics. The study of the structure of solutions to the general class of problems (P), with varying fidelity ϕ and regulariser R, is therefore of interest, as the analytical knowledge of properties of the solutions can provide vital insight and help in choosing the most suitable regularisation model for a given problem.

Starting with [28], and besides TGV and ICTV, various other higher-order regularisation schemes have been proposed in the recent years [8, 33, 12, 16, 17]. Curvature based regularisers such as Euler's elastica [14, 39] and [5] have also recently received attention for the better modelling of curvature in images. Further, non-convex total variation schemes are being studied for their better modelling of real image gradient distributions [26, 23, 24, 31, 25]. In the other direction, "lower-order schemes" such as Meyer's G-norm [29, 43] and TV with Kantorovich-Rubinstein discrepancy [27] have recently been proposed for the improved modelling of texture in images. Very little is known analytically about the solution of most of these models. We concentrate here on TV and, as a novelty, Huber-regularised and other TV variants with convex energies. We introduce these rigorously in §4. We also demonstrate that our technique would apply to non-convex TV schemes, as well as the Perona-Malik anisotropic diffusion [35, 44], if these were well-posed to begin with [25, 21]. In Part 2 we will discuss ICTV and TGV^2. We moreover hope that our techniques will be useful and provide an impetus for the analytical study of other advanced regularisers as well.

In case of TGV^2 regularisation, we know a little about solutions to (P) on one-dimensional domains $\Omega = (a,b)$ [7, 32]. In particular, we showed in [7] the jump set containment for $\phi(t) = t$, the proof easily extensible to $\phi(t) = t^2$. For first-order total variation regularisation, i.e. $R(u) = \alpha \mathrm{TV}(u) = \alpha \|Du\|_{\mathcal{M}(\Omega;\mathbb{R}^m)}$, the literature is more plentiful on analytical results. With the squared L^2 fidelity, the problem (P) is also known as the Rudin-Osher-Fatemi [38] problem (ROF), and written

$$\min_{u\in \mathrm{BV}(\Omega)} \frac{1}{2}\int_\Omega |f(x)-u(x)|^2\,dx + \alpha \mathrm{TV}(u). \tag{1.2}$$

The first structural results can be found in [37], where the stair-casing property is studied. Regarding the jump set, it is shown in [9] that $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$. The proofs of this and many other properties of total variation regularisation heavily depend on the co-area formula for functions of bounded variation, namely

$$\mathrm{TV}(u) = \int_{-\infty}^{\infty} \mathrm{Per}(\{u > t\};\Omega)\,dt, \tag{1.3}$$

where $\mathrm{Per}(E;\Omega) := \|D\chi_E\|_{2,\mathcal{M}(\Omega;\mathbb{R}^m)}$ denotes the perimeter of the set E within Ω. Thanks to the co-area formula, the problem (1.2) can be shown to be equivalent to the family of minimal surface problems

$$\min_{E\subset\Omega} \int_E (t - f(x))\,dx + \alpha\,\mathrm{Per}(E;\Omega), \quad (\text{a.e. } t \in \mathbb{R}). \tag{1.4}$$

The level sets of u can be found as solutions of (1.4) for varying t. This formulation and the regularity of the level sets is studied in [2, 10, 1].

For TV regularisation with the L^1 fidelity $\phi(x) = x$, i.e., the problem

$$\min_{u\in L^1(\Omega)} \int_\Omega |f(x)-u(x)|\,dx + \alpha \mathrm{TV}(u), \tag{1.5}$$

various basic properties were first studied by [13]. These include thresholds for the regularisation parameter $\alpha > 0$, under which the optimal solution u equals f. Similar thresholds were also derived in [7] for TGV^2 regularisation in the case $m = 1$. Some of these results readily generalise to $m > 1$. In this context of parameter thresholds, we also mention [30]. In [1] it is shown that the level sets $E_t = \{u \geq t\}$ of solutions u to (1.5), are solutions to the minimal surface problems

$$\min_{E\subset\Omega} \int_E \mathrm{sgn}(t - f(x))\,dx + \alpha\,\mathrm{Per}(E;\Omega), \quad (\text{a.e. } t \in \mathbb{R}). \tag{1.6}$$

Some further properties of the level sets are studied in [18]. We know from there that the essential inclusion $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$ does not generally hold, but the remainder has curvature α^{-1} for α the regularisation parameter. In a similar fashion, the capability of (1.5) to separate geometric objects according to their scales is studied in [45]. Finally, an interesting direction is taken in [4] by studying singular vectors and ground states of regularisation functionals, eventually hoping to obtain something resembling an eigendecomposition for solutions to (P). The singular vectors aside, all of these techniques heavily depend on the co-area formula.

As mentioned, in this paper, we introduce a novel technique for the study of the jump set J_u for rather general regularisers R, which is not based on the co-area formula. It is, instead, based on Lipschitz pushforwards. We push forward a purported solution by two different Lipschitz transformations and show that this provides a contradiction to u solving (P) if $\mathcal{H}^{m-1}(J_u \setminus J_f) > 0$, $(p > 1)$. Before starting to develop this technique in detail in §3, we now introduce some general notation, concepts and tools in the following §2.

2. Notation and useful facts. We begin by introducing the tools necessary for our work. First we introduce basic notation for sets, mappings, and measures. We then move on to Lipschitz mappings and graphs. Our notation and definition of the latter will be used extensively throughout the paper, as we perform operations on the jump set J_u, which is $\mathcal{H}^{m-1}$-rectifiable. Having defined our notation for Lipschitz graphs, we introduce functions of bounded variation.

2.1. Basic notations. We denote by $\{e_1, \ldots, e_m\}$ the standard basis of $\mathbb{R}^m$. The boundary of a set A we denote by ∂A, and the closure by $\operatorname{cl} A$. The $\{0, 1\}$-valued indicator function we write as χ_A. We denote the open ball of radius ρ centred at $x \in \mathbb{R}^m$ by $B(x, \rho)$. We denote by ω_m the volume of the unit ball $B(0, 1)$ in $\mathbb{R}^m$.

For $z \in \mathbb{R}^m$, we denote by $z^\perp := \{x \in \mathbb{R}^m \mid \langle z, x\rangle = 0\}$ the hyperplane orthogonal to z , whereas P_z denotes the projection operator onto the subspace spanned by z, and $P_z^\perp$ the projection onto $z^\perp$. If $A \subset z^\perp$, we denote by $\operatorname{ri} A$ the *relative interior* of A in $z^\perp$ as a subset of $\mathbb{R}^m$.

Let $\Omega \subset \mathbb{R}^m$ be an open set. We then denote the space of (signed) Radon measures on Ω by $\mathcal{M}(\Omega)$. If V is a vector space, then the space of Radon measures on Ω with values in V is denoted $\mathcal{M}(\Omega; V)$. The k-dimensional Hausdorff measure, on any given ambient space $\mathbb{R}^m$, $(k \le m)$, is denoted by $\mathcal{H}^k$, while $\mathcal{L}^m$ denotes the Lebesgue measure on $\mathbb{R}^m$.

The total variation (Radon) norm of a measure μ is denoted $\|\mu\|_{\mathcal{M}(\mathbb{R}^m)}$. For vector-valued measures $\mu = (\mu_1, \ldots, \mu^k) \in \mathcal{M}(\Omega; \mathbb{R}^k)$, we use the notation

$$\|\mu\|_{q,\mathcal{M}(\Omega;\mathbb{R}^k)} := \sup \left\{ \int_\Omega \sum_{i=1}^k \varphi_i(x)\, d\mu_i(x) \,\middle|\, \begin{array}{l} \varphi \in C_0^\infty(\Omega; \mathbb{R}^k), \\ \|\varphi(x)\|_p \le 1 \text{ for } x \in \Omega \end{array} \right\} \tag{2.1}$$

to indicate that the finite-dimensional base norm is the p-norm where $1/p + 1/q = 1$. When the choice of the finite-dimensional norm is inconsequential, we use the notation $\|\mu\|_{q,\mathcal{M}(\Omega;\mathbb{R}^k)}$. In this work in practise we restrict ourselves to $q = 2$ for measures. In other words, we consider isotropic total variation type functionals. We use the same notation for vector fields $w \in L^p(\Omega; \mathbb{R}^k)$, namely

$$\|w\|_{q,L^p(\Omega;\mathbb{R}^k)} := \left(\int_\Omega \|w(x)\|_q^p \, dx \right)^{1/p}.$$

For a measurable set A, we denote by $\mu \llcorner A$ the restricted measure defined by $(\mu \llcorner A)(B) := \mu(A \cap B)$. The notation $\mu \ll \nu$ means that μ is absolutely continuous with respect to the measure ν, and $\mu \perp \nu$ that μ and ν are mutually singular. The singular and absolutely continuous (with respect to the Lebesgue measure) part of μ are denoted μ^a and μ^s, respectively.

We denote the k-dimensional upper resp. lower density of μ by

$$\Theta_k^*(\mu; x) := \limsup_{\rho \searrow 0} \frac{\mu(B(x, \rho))}{\omega_k \rho^k}, \quad \text{resp.} \quad \Theta_{*,k}(\mu; x) := \liminf_{\rho \searrow 0} \frac{\mu(B(x, \rho))}{\omega_k \rho^k}.$$

The common value, if it exists, we denote by $\Theta_k(\mu; x)$.

Finally, we often denote by C, C', C''' arbitrary positive constants, and use the plus-minus notation $a^\pm = b^\pm$ in to mean that both $a^+ = b^+$ and $a^- = b^-$ hold.

2.2. Mappings from a subspace. We denote by $\mathcal{L}(V;W)$ the space of linear maps between the vector spaces V and W. If $L \in \mathcal{L}(V;\mathbb{R}^k)$, where $V \sim \mathbb{R}^n$, $(n \leq k)$, is a finite-dimensional Hilbert space, then $L^* \in \mathcal{L}(\mathbb{R}^k;V^*)$ denotes the adjoint, and the n-dimensional Jacobian is defined as [3]

$$\mathcal{J}_n[L] := \sqrt{\det(L^* \circ L)}.$$

With the gradient of a Lipschitz function $f : V \to \mathbb{R}^k$ defined in "components as columns order", $\nabla f(x) \in \mathcal{L}(\mathbb{R}^k;V)$, we extend this notation for brevity as

$$\mathcal{J}_n f(x) := \mathcal{J}_n[(\nabla f(x))^*].$$

If $\Omega \subset V$ is a Borel set, and $g \in L^1(\Omega)$, the *area formula* may then be stated

$$\int_{\mathbb{R}^k} \sum_{x \in \Omega \cap f^{-1}(y)} g(x)\, d\mathcal{H}^n(y) = \int_\Omega g(x)\mathcal{J}_n f(x)\, d\mathcal{H}^n(x). \tag{2.2}$$

That this indeed holds in our setting of finite-dimensional Hilbert spaces $V \sim \mathbb{R}^n$ follows by a simple argument from the area formula for $f : \mathbb{R}^n \to \mathbb{R}^k$, stated in, e.g, [3]. We only use the cases $V = z^\perp$ for some $z \in \mathbb{R}^m$ $(n = m-1)$, or $V = \mathbb{R}^m$ $(n = m)$.

We also denote by

$$C^{2,\cap}(V) := \bigcap_{\lambda \in (0,1)} C^{2,\lambda}(V)$$

the class of functions that are twice differentiable (as defined above for tensor fields) with a Hölder continuous second differential for all exponents $\lambda \in (0,1)$.

The Lipschitz factor of a Lipschitz mapping f we denote by $\operatorname{lip} f$. We also recall that a Lipschitz transformation $\gamma : U \to V$ with $U, V \subset \mathbb{R}^m$ has the *Lusin N-property* if it maps $\mathcal{L}^m$-negligible sets to $\mathcal{L}^m$-negligible sets.

If $\gamma : \Omega \to \Omega$ is a bijective Lipschitz transformation, and $u : \Omega \to \Omega$ a Borel function, we define the pushforward $u_\gamma := \gamma_\# u := u \circ \gamma^{-1}$. Finally, we denote the identity transformation by $\iota(x) = x$.

2.3. Lipschitz graphs. A set $\Gamma \subset \mathbb{R}^m$ is called a Lipschitz $(m-1)$-graph (of Lipschitz factor L), if there exist a unit vector z_Γ, an open set $V_\Gamma \subset z_\Gamma^\perp$, and a Lipschitz map $f_\Gamma : V_\Gamma \to \mathbb{R}$, of Lipschitz factor at most L, such that

$$\Gamma = \{v + f_\Gamma(v) z_\Gamma \mid v \in V_\Gamma\}.$$

We also define $g_\Gamma : V_\Gamma \to \mathbb{R}^m$ by

$$g_\Gamma(v) = v + z_\Gamma f_\Gamma(v).$$

Then

$$\Gamma = g_\Gamma(V_\Gamma).$$

We denote the open domains "above" and "beneath" Γ, respectively, by

$$\Gamma^+ := \Gamma + (0,\infty) z_\Gamma, \quad \text{and} \quad \Gamma^- := \Gamma + (-\infty, 0) z_\Gamma.$$

We recall that by Kirszbraun's theorem, we may extend the domain of f_Γ and g_Γ from V_Γ to the whole space $z_\Gamma^\perp$ without altering the Lipschitz constant. Then Γ splits Ω into the two open halves $\Gamma^+ \cap \Omega$ and $\Gamma^- \cap \Omega$. We often use this fact.

2.4. Functions of bounded variation. We say that a function $u : \Omega \to \mathbb{R}$ on a bounded open set $\Omega \subset \mathbb{R}^m$ is of *bounded variation* (see, e.g., [3] for a more thorough introduction), denoted $u \in \mathrm{BV}(\Omega)$, if $u \in L^1(\Omega)$, and the distributional gradient Du is a Radon measure. Given a sequence $\{u^i\}_{i=1}^\infty \subset \mathrm{BV}(\Omega)$, weak* convergence is defined as $u^i \to u$ strongly in $L^1(\Omega)$ along with $Du^i \overset{*}{\rightharpoonup} Du$ weakly* in $\mathcal{M}(\Omega)$. The sequence converges *strictly* if, in addition to this, $|Du^i|(\Omega) \to |Du|(\Omega)$.

We denote by S_u the approximate discontinuity set, i.e., the complement of the set where the Lebesgue limit $\widetilde{u}$ exists. The latter is defined by

$$\lim_{\rho \searrow 0} \frac{1}{\rho^m} \int_{B(x,\rho)} \|\widetilde{u}(x) - u(y)\|\, dy = 0.$$

The distributional gradient can be decomposed as $Du = \nabla u \mathcal{L}^m + D^j u + D^c u$, where the density ∇u of the *absolutely continuous part* of Du equals (a.e.) the approximate differential of u. We also define the *singular part* as $D^s u = D^j u + D^c u$. The *jump part* $D^j u$ may be represented as

$$D^j u = (u^+ - u^-) \otimes \nu_{J_u} \mathcal{H}^{m-1} \llcorner J_u,$$

where x is in the *jump set* $J_u \subset S_u$ of u if for some $\nu := \nu_{J_u}(x)$ there exist two *distinct* one-sided traces $u^+(x)$ and $u^-(x)$, defined as satisfying

$$\lim_{\rho \searrow 0} \frac{1}{\rho^m} \int_{B^\pm(x,\rho,\nu)} \|u^\pm(x) - u(y)\|\, dy = 0,$$

where $B^\pm(x,\rho,\nu) := \{y \in B(x,\rho) \mid \pm\langle y - x, \nu\rangle \geq 0\}$. It turns out that J_u is countably $\mathcal{H}^{m-1}$-rectifiable and ν is (a.e.) the normal to J_u. This former means that there exist Lipschitz $(m-1)$-graphs $\{\Gamma_i\}_{i=1}^\infty$ such that $\mathcal{H}^{m-1}(J_u \setminus \bigcup_{i=1}^\infty \Gamma_i) = 0$. Moreover, we have $\mathcal{H}^{m-1}(S_u \setminus J_u) = 0$. The remaining *Cantor part* $D^c u$ vanishes on any Borel set σ-finite with respect to $\mathcal{H}^{m-1}$.

We will frequently use the following basic properties of densities of Du; for the proof, see, e.g., [3, Proposition 3.92].

PROPOSITION 2.1. *Let $u \in \mathrm{BV}(\Omega)$ for an open domain $\Omega \subset \mathbb{R}^m$. Define*

$$\widetilde{S}_u := \{x \in \Omega \mid \Theta_{*,m}(|Du|; x) = \infty\}, \quad \text{and} \quad \widetilde{J}_u := \{x \in \Omega \mid \Theta_{*,m-1}(|Du|; x) > 0\}. \tag{2.3}$$

Then the following decomposition holds.

(i) $\nabla u \mathcal{L}^m = Du \llcorner (\Omega \setminus \widetilde{S}_u)$.

(ii) $D^j u = Du \llcorner \widetilde{J}_u$, *precisely* $\widetilde{J}_u \supset J_u$, *and* $\mathcal{H}^{m-1}(\widetilde{J}_u \setminus J_u) = 0$.

(iii) $D^c u = Du \llcorner (\widetilde{S}_u \setminus \widetilde{J}_u)$.

3. Problem statement. We now have most of the tools needed to state our main results, particularly the containment of J_u in J_f modulo a $\mathcal{H}^{m-1}$-null set for convex p-increasing $(p > 1)$ ϕ. We just have to rigorously state our assumptions on the regularisation functional R and the fidelity ϕ. These ensure firstly the existence of a solution $u \in \mathrm{BV}(\Omega)$ to (P). Secondly we want to ensure that R behaves almost like TV under averaged Lipschitz transformations, and ϕ comparably slower.

3.1. Admissible regularisation functionals and fidelities. We begin by stating our assumptions on R.

DEFINITION 3.1. *We call R an* admissible regularisation functional *on $L^1(\Omega)$, where the domain $\Omega \subset \mathbb{R}^m$, if it is convex, lower semi-continuous with respect to* *weak* convergence in* $\mathrm{BV}(\Omega)$*, and there exists* $C > 0$ *such that*

$$\|Du\|_{\mathcal{M}(\Omega;\mathbb{R}^m)} \le C\bigl(1 + \|u\|_{L^1(\Omega)} + R(u)\bigr), \quad (u \in L^1(\Omega)). \tag{3.1}$$

Here and throughout the paper, unless otherwise indicated, the finite-dimensional base norm on $\mathbb{R}^m$ *is the Euclidean or 2-norm. This is the one most appropriate to most image processing tasks due to its rotational invariance.*

DEFINITION 3.2. *We denote by* $\mathcal{F}(\Omega)$ *the set of one-to-one Lipschitz transformations* $\gamma : \Omega \to \Omega$ *with* γ^{-1} *also Lipschitz and both satisfying the Lusin* N*-property. With* $U \subset \Omega$ *an open set, we further denote*

$$\mathcal{F}(\Omega, U) := \{\gamma \in \mathcal{F}(\Omega) \mid \gamma(x) = x \text{ for } x \notin U\}.$$

With $\overline{\gamma}, \underline{\gamma} \in \mathcal{F}(\Omega)$*, we then define the* basic double-Lipschitz comparison constants

$$G_{\overline{\gamma},\underline{\gamma}} := \sup_{x\in\Omega, v\in\mathbb{R}^m, \|v\|=1} \|A_{\overline{\gamma}}(x)v\| + \|A_{\underline{\gamma}}(x)v\| - 2\|v\|. \tag{3.2a}$$

and

$$\bar{J}_{\overline{\gamma},\underline{\gamma}} := \sup_{x\in\Omega} |\mathcal{J}_m\overline{\gamma}(x) + \mathcal{J}_m\underline{\gamma}(x) - 2|. \tag{3.2b}$$

Here I *is the identity mapping on* $\mathbb{R}^m$*, and*

$$A_\gamma(x) := \nabla\gamma^{-1}(\gamma(x))\mathcal{J}_m\gamma(x). \tag{3.2c}$$

We also define the distance-to-identity

$$D_\gamma := \sup_{x\in\Omega} \|\nabla\gamma^{-1}(\gamma(x)) - I\|, \tag{3.2d}$$

and finally combine all of these into the overall double-Lipschitz comparison constant

$$T_{\overline{\gamma},\underline{\gamma}} := G_{\overline{\gamma},\underline{\gamma}} + \bar{J}_{\overline{\gamma},\underline{\gamma}} + D_{\overline{\gamma}}^2 + D_{\underline{\gamma}}^2. \tag{3.2e}$$

Remark 3.1. With the help of $\bar{J}_{\overline{\gamma},\underline{\gamma}}$ and a bound on the Lipschitz factors, we could generally replace $G_{\overline{\gamma},\underline{\gamma}}$ by

$$\widetilde{G}_{\overline{\gamma},\underline{\gamma}} := \sup_{x\in\Omega, \|v\|=1} \|\nabla\overline{\gamma}^{-1}(\gamma(x))v\| + \|\nabla\underline{\gamma}^{-1}(\gamma(x))v\| - 2\|v\|.$$

It however turns out that for our transformations of interest, A_γ is easier work with than $\nabla\gamma^{-1} \circ \gamma$ directly. With the help of a bound on the Lipschitz factor, we could also replace D_γ by $G_{\gamma,\iota}$, and then $T_{\overline{\gamma},\underline{\gamma}}$ by

$$\widetilde{T}_{\overline{\gamma},\underline{\gamma}} := \overline{T}_{\overline{\gamma},\underline{\gamma}} + \overline{T}_{\overline{\gamma},\iota}^2 + \overline{T}_{\underline{\gamma},\iota}^2,$$

where $\overline{T}_{\overline{\gamma},\underline{\gamma}} := G_{\overline{\gamma},\underline{\gamma}} + \bar{J}_{\overline{\gamma},\underline{\gamma}}$ provides a form of simultaneous distance of the two transformations to the identity. Various further over-estimations of $T_{\overline{\gamma},\underline{\gamma}}$ are possible, but these usually destroy the crucial $O(\rho^2)$ property that we will later derive for specific transformations. At this point, it is threfore not clear whether $T_{\overline{\gamma},\underline{\gamma}}$ in general can be replaced by something significantly simpler and intuitive; alternative uses of the comparison condition with different transformations from those in the present work, may have different requirements from the useful set of comparison constants, possibly in some ways sharper than those herein.

DEFINITION 3.3. *We say that R is* double-Lipschitz comparable *if there exists a constant $R^a = R^a(\Omega)$ such that*

$$\text{for every } \begin{cases} u \in \mathrm{BV}(\Omega), \\ \text{open set } U \subset \Omega, \text{ and} \\ \overline{\gamma}, \underline{\gamma} \in \mathcal{F}(\Omega, U) \text{ with } T_{\overline{\gamma},\underline{\gamma}} < 1, \end{cases}$$

there holds

$$R(\overline{\gamma}_{\#}u) + R(\underline{\gamma}_{\#}u) - 2R(u) \le R^a T_{\overline{\gamma},\underline{\gamma}} |Du|(\operatorname{cl} U).$$

We also say that R is separably double-Lipschitz comparable *if there exist constants $R^a = R^a(\Omega)$ and $R^s = R^s(\Omega)$ such that*

$$\text{for every } \begin{cases} u \in \mathrm{BV}(\Omega), \\ \text{open set } U \subset \Omega, \\ \overline{\gamma}, \underline{\gamma} \in \mathcal{F}(\Omega, U) \text{ with } T_{\overline{\gamma},\underline{\gamma}} < 1, \quad \text{and} \\ \text{Lipschitz } (m-1)\text{-graph } \Gamma \end{cases}$$

holds

$$\begin{aligned} R(\overline{\gamma}_{\#}u) + R(\underline{\gamma}_{\#}u) - 2R(u) &\le R^a T_{\overline{\gamma},\underline{\gamma}} |Du|(\operatorname{cl} U \setminus \Gamma) \\ &\quad + R^s \big(|D\overline{\gamma}_{\#}u|(\overline{\gamma}(\Gamma)) + |D\underline{\gamma}_{\#}u|(\underline{\gamma}(\Gamma)) - 2|Du|(\Gamma) \big). \end{aligned}$$

Remark 3.2. Strictly speaking, we only need a local $\mathcal{H}^{m-1}$-a.e. version of double-Lipschitz comparability, but the regularisation functionals in this Part 1 satisfy the stronger and simpler definition above. Therefore we use it. In Part 2, we will need to consider much more detailed variants. Also the bound $T_{\overline{\gamma},\underline{\gamma}} < 1$ is primarily needed for $D_{\overline{\gamma}}, D_{\underline{\gamma}} < 1$ and an arbitrary bound on $\bar{J}_{\overline{\gamma},\underline{\gamma}}$ for the treatment of Huber-regularised TV.

In order to show the existence of solutions to (P), we require the following property from ϕ.

DEFINITION 3.4. *Let the domain $\Omega \subset \mathbb{R}^m$. We call $\phi : \mathbb{R} \to [0, \infty]$ an* admissible fidelity function on Ω *if ϕ is convex and lower semi-continuous, $\phi(0) = 0$, $\phi(-t) = \phi(t)$, $(t > 0)$, and for some $C > 0$ holds*

$$\|u\|_{L^1(\Omega)} \le C \left(\int_\Omega \phi(u(x))\, dx + 1 \right), \quad (u \in L^1(\Omega)). \tag{3.3}$$

For the study of the jump set J_u of solutions to (P), we require additionally the following increase criterion to be satisfied by ϕ.

DEFINITION 3.5. *We say that ϕ is p-*increasing *for $p \ge 1$, if there exists a constant $C_\phi > 0$ for which*

$$\phi(x) - \phi(y) \le C_\phi(|x| - |y|)|x|^{p-1}, \quad (x, y \in \mathbb{R}).$$

Remark 3.3. Definition 3.5 implies that ϕ is increasing. In fact $\phi'(x) \geq C_\phi |x|^{p-1}$.

Example 3.1. Let $\phi(x) = |x|^p$, $(p \geq 1)$. Then ϕ is p-increasing with $C_\phi = p$ because

$$\phi(x) - \phi(y) \leq \phi'(x)(|x| - |y|) = p(|x| - |y|)|x|^{p-1}.$$

The L^1 fidelity $\phi(x) = |x|$ is admissible for any $\Omega \subset \mathbb{R}^m$, including $\Omega = \mathbb{R}^m$. The L^p fidelities $\phi(x) = x^p$ for $p > 1$ are admissible for any bounded $\Omega \subset \mathbb{R}^m$.

The following, standard, result states that the problem (P) is well-posed under the above assumptions. We may therefore proceed with the analysis of the structure of the solutions $u \in \mathrm{BV}(\Omega)$.

THEOREM 3.6. *Let $f \in L^1(\Omega)$ satisfy $\int_\Omega \phi(f(x))\,dx < \infty$. Suppose that R is an admissible regularisation functional on $L^1(\Omega)$, and ϕ an admissible fidelity function for Ω. Then there exists a solution $u \in L^1(\Omega)$ to (P), and any solution satisfies $u \in \mathrm{BV}(\Omega)$.*

Proof. Clearly the minimum in (P) is finite. Let $\{u^i\}_{i=0}^\infty$ be a minimising sequence for (P). Minding that (3.1), (3.3) bound $\{u^i\}_{i=0}^\infty$ in $\mathrm{BV}(\Omega)$, and that R is lower semi-continuous with respect to weak* convergence in $\mathrm{BV}(\Omega)$, and ϕ is lower semi-continuous, the claim follows by the standard method of calculus of variations, see, e.g., [19]. □

3.2. The main results. Our main task in the rest of this paper is to prove the following result.

THEOREM 3.7. *Let the domain $\Omega \subset \mathbb{R}^m$ be bounded with Lipschitz boundary. Suppose R is an admissible double-Lipschitz comparable regularisation functional on $L^1(\Omega)$, and $\phi : [0, \infty) \to [0, \infty)$ an admissible p-increasing fidelity function for some $1 < p < \infty$. Let $f \in \mathrm{BV}(\Omega) \cap L^\infty_{\mathrm{loc}}(\Omega)$, and suppose $u \in \mathrm{BV}(\Omega) \cap L^\infty_{\mathrm{loc}}(\Omega)$ solves (P). Then*

$$\mathcal{H}^{m-1}(J_u \setminus J_f) = 0.$$

Remark 3.4. Observe that we require u to be locally bounded. This does not necessarily hold, and needs to be proved separately. It is well-known that it holds for TV if $f \in L^\infty(\Omega)$, and is easy to show for Huber-TV using the same barrier technique.

We also show the following. We note that we only get strong regularity if the solution is approximately piecewise constant.

THEOREM 3.8. *Let the domain $\Omega \subset \mathbb{R}^m$ be bounded with Lipschitz boundary. Suppose R is an admissible separably double-Lipschitz comparable regularisation functional on $L^1(\Omega)$, and $\phi : [0, \infty) \to [0, \infty)$ an admissible 1-increasing fidelity function. Let $f \in \mathrm{BV}(\Omega)$, and suppose $u \in \mathrm{BV}(\Omega)$ solves (P). Then*

$$\mathcal{H}^{m-1}(J_u \setminus (J_f \cup \Lambda)) = 0,$$

where $\Lambda = \bigcup_{i=1}^\infty \Lambda_i$ for Lipschitz graphs $\{\Lambda_i\}_{i=1}^\infty$ such that at $\mathcal{H}^{m-1}$-a.e. point $x \in (\Lambda_i \cap J_u) \setminus J_f$, the R-curvature of u along Λ_i *satisfies*

$$\mathcal{C}_u^{R,\Lambda_i}(x) = |u^+(x) - u^-(x)| C_\phi. \tag{3.4}$$

This technical definition will be provided later in Definition 7.1.

If $\Theta_m(|Du|\llcorner\Omega \setminus \Lambda_i; x) = 0$ at $\mathcal{H}^{m-1}$-a.e. point $x \in \Lambda_i$, then each Λ_i, $(i \in \mathbb{Z}^+)$, is of class $C^{2,\cap}$ and curvature C_ϕ/R^s in the sense that $f_{\Lambda_i} \in C^{2,\cap}(V_{\Lambda_i})$ and

$$-\operatorname{div} \frac{\nabla f_{\Lambda_i}(v)}{\sqrt{1 + \|\nabla f_{\Lambda_i}(v)\|^2}} = C_\phi/R^s, \quad (v \in V_{\Lambda_i}). \tag{3.5}$$

Remark 3.5. It is not very difficult to improve Theorem 3.8 a little bit. Namely, we can replace the assumption $\Theta_m(|Du|\llcorner\Omega \setminus \Lambda_i; x) = 0$ by that of ∇u having one-sided Lebesgue limits at x, with corresponding normal $\nu = \nu_{J_u}(x)$. We will in another context in Part 2 study techniques that would allow us to do this. We do not however pursue this route of improving Theorem 3.8, as the small improvement would still not be entirely satisfactory – at least not without corresponding results to prove that the limits actually do exist. This fascinating question is outside the scope of the present manuscript.

Remark 3.6. For a demonstration that we cannot in general set $\Lambda = \emptyset$ in Theorem 3.8, we refer to the comprehensive treatment in [18] about the case of $R = \mathrm{TV}$, $\phi = |\cdot|$, and $f = \chi_A$ for a suitable set $A \subset \Omega$. For example if A is a square, then the optimal solution u will be a square with rounded corners of curvature $1/\alpha = C_\phi/R^s$ (or the empty set, when this is not possible). Using [34, Theorem 8], this example can also be extended to the higher-order regulariser TGV^2 [6], which we treat in Part 2 [41].

4. First-order regularisation functionals. Before embarking on the proofs of the main results, we introduce a class of admissible first-order regularisation functionals: the conventional total variation, as well as a class with an additional convex energy, including Huber-regularised TV. We finish by discussing the Perona-Malik and non-convex TV models in a few remarks. In Part 2 [41] we will concentrate on higher-order regularisation functionals: second-order total generalised variation (TGV^2), and infimal convolution TV (ICTV), whose analysis is more involved. We however remark that both Theorem 3.7 and Theorem 3.8 can easily be derived for ICTV from the corresponding result for TV. For TGV^2 this is not the case.

4.1. Total variation. As we well recall, (isotropic) total variation is defined as

$$\mathrm{TV}(u) := |Du|(\Omega) = \sup \left\{ \int_\Omega \langle \operatorname{div} \varphi(x), u(x) \rangle \, dx \,\middle|\, \varphi \in C_c^\infty(\Omega), \|\varphi\|_{2,L^\infty(\Omega)} \leq 1 \right\}.$$

We now show that it satisfies the conditions of Theorem 3.7.

PROPOSITION 4.1. *The functional $R(u) = \alpha \mathrm{TV}(u)$ for $\alpha > 0$ is admissible and (separably) double-Lipschitz comparable. Moreover, $\|u\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)}$ for solutions u to (P).*

We use the following simple lemma for the proof.

LEMMA 4.2. *Let $u \in \mathrm{BV}(\Omega)$ and $\gamma \in \mathcal{F}(\Omega)$. Then $D^a(\gamma_\# u) = \nabla\gamma^{-1}\gamma_\#\nabla u \mathcal{L}^m$, that is,*

$$\nabla(\gamma_\# u)(x) = \nabla\gamma^{-1}(x)\gamma_\#\nabla u(x) = \gamma_\#([\nabla\gamma]^{-1}\nabla u)(x), \quad (\mathcal{L}^m\text{-a.e. } x \in \Omega).$$

Proof. We observe first of all that the inverse function theorem trivially holds almost everywhere for $\gamma \in \mathcal{F}(\Omega)$. Indeed, using Rademacher's theorem and the Lusin N-property, we see that $\iota(x) = \gamma(\gamma^{-1}(x))$ satisfies $\nabla\iota(x) = \nabla\gamma(\gamma^{-1}(x))\nabla\gamma^{-1}(x)$ for

$\mathcal{L}^m$-a.e. $x \in \Omega$. Repeating the same argument on the formulation $\iota(x) = \gamma^{-1}(\gamma(x))$, therefore $[\nabla\gamma(\gamma^{-1}(x))]^{-1} = \nabla\gamma^{-1}(x)$ for $\mathcal{L}^m$-a.e. $x \in \Omega$.

By the Calderón-Zygmund theorem, $D^a v = \nabla v \mathcal{L}^m$, where ∇v is the approximate differential of $v \in \mathrm{BV}(\Omega)$. This is defined at almost every $x \in \Omega$ as $\nabla v(x) = L$, where L satisfies

$$\lim_{\rho \searrow} \fint_{B(x,\rho)} \frac{|v(y) - \widetilde{v}(x) - \langle L, y-x\rangle|}{\rho}\, dy = 0.$$

Let $x \in \Omega$ be a point such that $\widetilde{u}(y)$ and $\nabla u(y)$ exist for $y = \gamma^{-1}(x)$. Since γ has the Lusin N-property, $\mathcal{L}^m$-a.e. $x \in \Omega$ satisfies this. Clearly, by a simple application of the area formula, if $v = \gamma_\# u$, then $\widetilde{v}(x) = \widetilde{u}(y)$. Therefore, if we define $L = [\nabla\gamma(y)]^{-1}\nabla u(y) = \nabla\gamma^{-1}(x)\gamma_\#\nabla u(x)$, it is easily seen that $\nabla(\gamma_\# u)(x) = L$ exists, and has the required form. □

Proof of Proposition 4.1. The requirements of admissibility in Definition 3.1 are trivial in this case. Also $\|u\|_{L^\infty(\Omega)} \le M := \|f\|_{L^\infty(\Omega)}$ is well-known for solutions u to (P) with total variation regularisation. Indeed, comparing a purported solution u that violates this with $\widetilde{u} := \min\{\max\{-M, u\}, M\}$, and referring to the co-area formula (1.3), we easily obtain a contradiction.

We therefore only have to prove (separable) double-Lipschitz comparability. We may without loss of generality take $\alpha = 1$. We let $\gamma \in \mathcal{F}(\Omega; U)$ for some open set $U \subset \Omega$, and pick $u \in \mathrm{BV}(\Omega)$. By Lemma 4.2, we have

$$\begin{aligned} |D\gamma_\# u|(\Omega) &= |D^a\gamma_\# u|(\Omega) + |D^s\gamma_\# u|(\Omega) \\ &= |\nabla\gamma^{-1}\gamma_\#\nabla u\mathcal{L}^m|(\Omega) + |D^s\gamma_\# u|(\Omega). \end{aligned} \tag{4.1}$$

Since $\gamma(x) = x$ for $x \in \Omega \setminus U$, we may calculate using the area formula

$$\begin{aligned} |\nabla\gamma^{-1}\gamma_\#\nabla u\mathcal{L}^m|(\Omega) &= \int_\Omega \|\nabla\gamma^{-1}(x)\nabla u(\gamma^{-1}(x))\|\, dx \\ &= |D^a u|(\Omega \setminus U) + \int_U \|A_\gamma(x)\nabla u(x)\|\, dx. \end{aligned}$$

Thus, with $\overline{\gamma}, \underline{\gamma} \in \mathcal{F}(\Omega, U)$, referring to the definition of $G_{\overline{\gamma},\underline{\gamma}}$, we obtain

$$|\nabla\overline{\gamma}^{-1}\overline{\gamma}_\#\nabla u\mathcal{L}^m|(\Omega) + |\nabla\underline{\gamma}^{-1}\underline{\gamma}_\#\nabla u\mathcal{L}^m|(\Omega) - 2|D^a u|(\Omega) \le G_{\overline{\gamma},\underline{\gamma}}|D^a u|(U).$$

Recalling (4.1), and minding that

$$|D^s\overline{\gamma}_\# u|(\Omega \setminus U) = |D^s\underline{\gamma}_\# u|(\Omega \setminus U) = |D^s u|(\Omega \setminus U),$$

we deduce

$$\begin{aligned} &|D\overline{\gamma}_\# u|(\Omega) + |D\underline{\gamma}_\# u|(\Omega) - 2|Du|(\Omega) \\ &\qquad \le G_{\overline{\gamma},\underline{\gamma}}|D^a u|(U) + \big(|D^s\overline{\gamma}_\# u|(U) + |D^s\underline{\gamma}_\# u|(U) - 2|D^s u|(U)\big). \end{aligned} \tag{4.2}$$

This almost proves (separable) double-Lipschitz comparability, we just have to modify the singular part appropriately.

To see (non-separable) double-Lipschitz comparability, we take a strictly converging approximation sequence $\{u^i\}_{i=1}^\infty \subset C^1(\Omega)$ of u. Then $D^s u^i = 0$, so (4.2) proves

$$|D\overline{\gamma}_\# u^i|(\Omega) + |D\underline{\gamma}_\# u^i|(\Omega) - 2|Du^i|(\Omega) \le G_{\overline{\gamma},\underline{\gamma}}|D^a u^i|(U), \quad (i = 1, 2, 3, \ldots).$$

Since the strict convergence of $\{u^i\}_{i=1}^\infty$ to u bounds the right hand side, we deduce

$$\sup_i |D\overline{\gamma}_\# u^i|(\Omega) + \sup_i |D\underline{\gamma}_\# u^i|(\Omega) < \infty.$$

We may therefore extract a subsequence, unrelabelled, such that both $\{\overline{\gamma}_\# u^i\}_{i=1}^\infty$ and $\{\underline{\gamma}_\# u^i\}_{i=1}^\infty$ are convergent weakly* to some $\overline{u} \in \mathrm{BV}(\Omega)$ and $\underline{u} \in \mathrm{BV}(\Omega)$, respectively. Moreover, by lower semicontinuity of the total variation, and strict convergence of the approximating sequence

$$\begin{aligned}|D\overline{u}|(\Omega) + |D\underline{u}|(\Omega) - 2|Du|(\Omega) &\leq \liminf_{i\to\infty} |D\overline{\gamma}_\# u^i|(\Omega) + |D\underline{\gamma}_\# u^i|(\Omega) - 2|Du^i|(\Omega) \\ &\leq \liminf_{i\to\infty} G_{\overline{\gamma},\underline{\gamma}}|Du^i|(U).\end{aligned}$$

Let us pick an open set $U' \supset U$ such that $|Du|(\partial U') = 0$. Then $|Du^i|(U') \to |Du|(U')$ because $u^i \to u$ strictly in $L^1(\Omega)$; see [3, Proposition 1.62]. It follows

$$|D\overline{u}|(\Omega) + |D\underline{u}|(\Omega) - 2|Du|(\Omega) \leq G_{\overline{\gamma},\underline{\gamma}}|Du|(U').$$

By taking the intersection over all admissible $U' \supset U$, we deduce

$$|D\overline{u}|(\Omega) + |D\underline{u}|(\Omega) - 2|Du|(\Omega) \leq G_{\overline{\gamma},\underline{\gamma}}|Du|(\operatorname{cl} U). \tag{4.3}$$

This is almost the double-Lipschitz comparability. We just have to show that $\overline{u} = \overline{\gamma}_\# u$ and $\underline{u} = \underline{\gamma}_\# u$. Indeed

$$\begin{aligned}\int_\Omega |\overline{u}(x) - \overline{\gamma}_\# u(x)|\,dx &\leq \int_\Omega |\overline{u}(x) - \overline{\gamma}_\# u^i(x)|\,dx + \int_\Omega |\overline{\gamma}_\# u(x) - \overline{\gamma}_\# u^i(x)|\,dx \\ &\leq \int_\Omega |\overline{u}(x) - \overline{\gamma}_\# u^i(x)|\,dx + C\int_\Omega |u(x) - u^i(x)|\,dx,\end{aligned}$$

where

$$C := \sup_x \mathcal{J}_m \gamma(x) \leq (\operatorname{lip}\gamma)^m < \infty.$$

But the integrals on the right hand side tend to zero since the strict and weak* convergences imply strong convergence in L^1 of $\{u^i\}_{i=1}^\infty$ to u and of $\{\overline{\gamma}_\# u^i\}_{i=1}^\infty$ to $\overline{u}$. It follows that $\overline{u} = \overline{\gamma}_\# u$. Analogously we show that $\underline{u} = \underline{\gamma}_\# u$. Double-Lipschitz comparability is now immediate from (4.3).

Finally, to see separable double-Lipschitz comparability, we proceed analogously as above, but approximate u on both sides of Γ. More specifically, referring to Kirzbraun's theorem, we may assume that $V_\Gamma = z^\perp$. Thus Ω splits into two domains $\Omega^\pm := \Omega \cap \Gamma^\pm$. We approximate u separately on both Ω^+ and Ω^- by strictly converging sequences of C^1 functions $\{u^{(+),i}\}_{i=1}^\infty$ and $\{u^{(-),i}\}_{i=1}^\infty$. By the continuity of the trace operator with respect to strict convergence in $\mathrm{BV}(\Omega)$ (see, e.g., [3, Theorem 3.88]), also $u^i := u^{(+),i} + u^{(-),i}$ then converge strictly to u. Moreover, $D^s u^i = Du^i \llcorner \Gamma$. By (4.2) we therefore have

$$\begin{aligned}&|D\overline{\gamma}_\# u^i|(\Omega) + |D\underline{\gamma}_\# u^i|(\Omega) - 2|Du^i|(\Omega) \\ &\qquad \leq G_{\overline{\gamma},\underline{\gamma}}|D^a u^i|(U) + \big(|D\overline{\gamma}_\# u^i|(\overline{\gamma}(\Gamma)) + |D\underline{\gamma}_\# u^i|(\underline{\gamma}(\Gamma)) - 2|Du^i|(\Gamma)\big).\end{aligned}$$

Since the traces of u^i on Γ converge in $L^1(\Gamma)$ to the trace of u on Γ, we deduce the separable double-Lipschitz property by analogous arguments as the (non-separable) double-Lipschitz property above. □

Remark 4.1. Observe that we obtained the estimate

$$\mathrm{TV}(\overline{\gamma}_{\#}u) + \mathrm{TV}(\underline{\gamma}_{\#}u) - 2\mathrm{TV}(u) \le G_{\overline{\gamma},\underline{\gamma}}|Du|(\mathrm{cl}\, U)$$

that involves none of $D^2_{\overline{\gamma}}$, $D^2_{\underline{\gamma}}$, and $\bar{J}_{\overline{\gamma},\underline{\gamma}}$. If we take $\underline{\gamma}(x) = \iota(x) := x$, then the property shows

$$|D\overline{\gamma}_{\#}u|(\Omega) \le (1 + G_{\overline{\gamma},\iota})\, |Du|(\Omega).$$

This can in specific cases can improve upon standard estimates [3] on the total variation under Lipschitz transformations.

Observe also that our proof does not depend on the finite-dimensional norm in the definition of TV being the Euclidean norm. The result holds for any choice of finite-dimensional norm as long as the definition of $G_{\overline{\gamma},\underline{\gamma}}$ also reflects this. In §5 we will however depend on the properties of the Euclidean norm.

4.2. Huber-regularised TV. For a parameter $\eta > 0$, Huber-regularised total variation may be defined as

$$\mathrm{TV}_\eta(u) := \sup\left\{ \int_\Omega u \,\mathrm{div}\, \varphi \, dx - \frac{\eta}{2}\|\varphi\|^2_{L^2(\Omega;\mathbb{R}^m)} \;\middle|\; \begin{array}{l} \varphi \in C_c^\infty(\Omega;\mathbb{R}^m), \\ \|\varphi\|_{2,L^\infty(\Omega;\mathbb{R}^m)} \le 1 \end{array} \right\}. \tag{4.4}$$

If u is smooth, this corresponds to replacing the pointwise 2-norm by

$$|g|_\eta = \begin{cases} \|g\|_2 - \frac{1}{2\eta}, & \|g\|_2 \ge 1/\eta, \\ \frac{\eta}{2}\|g\|_2^2, & \|g\|_2 < 1/\eta, \end{cases}$$

giving

$$\mathrm{TV}_\eta(u) = \int_\Omega |\nabla u(x)|_\eta \, dx, \quad (C^1(\Omega)).$$

Huber-regularisation of TV is sometimes helpful numerically, especially in the context of second-order optimisation methods [22], as well as primal-dual methods for non-convex problems [42]. It also helps to avoid the stair-casing effect to some extent. As with higher-order regularisers, there is no apparent useful coarea formula for Huber-TV, that would allow us to show regularity properties through level sets. However, TV_η satisfies our assumptions, as stated in the following. In fact, since our proof is based on rather general properties, we consider a slightly larger class of functionals, based on a class of convex energies.

Definition 4.3. *We denote by Ψ the family of increasing convex functions $\psi : [0,\infty) \to [0,\infty)$ with $\psi(0) = 0$, that satisfy the following two properties. Firstly, $0 < \psi^\infty < \infty$ for*

$$\psi^\infty := \lim_{t \nearrow \infty} \psi(t)/t.$$

Secondly, ψ satisfies for some constants $K_\psi, C_\psi > 0$ the property

$$\langle \partial\psi(t) - \partial\psi(s), t - s\rangle \le \begin{cases} 0, & \min\{t,s\} \in [K_\psi, \infty), \\ C_\psi|t-s|^2, & \min\{t,s\} \in [0, K_\psi). \end{cases} \tag{4.5}$$

In essence, the threshold K_ψ says that TV_ψ, defined after the next lemma, behaves linearly like TV for large gradients. This of course holds for Huber-TV.

Lemma 4.4. *We have $|\cdot|_\eta \in \Psi$, $(\eta > 0)$.*

Proof. Only (4.5) demands proof. We set $\psi(t) := |t|_\eta$. Clearly, if $s, t \geq 1/\eta$, we have

$$\langle \partial\psi(t) - \partial\psi(s), t - s\rangle = 0,$$

so the property holds. If $s, t < 1/\eta$, we have

$$\langle \partial\psi(t) - \partial\psi(s), t - s\rangle = \eta|t - s|^2,$$

so we have (4.5) in this range. In the case $s < 1/\eta < t$, we also have

$$\langle \partial\psi(t) - \partial\psi(s), t - s\rangle = (1 - \eta s)(t - s) \leq (\eta t - \eta s)(t - s) \leq \eta|t - s|^2.$$

The property (4.5) follows with $K_\psi = 1/\eta$ and $C_\psi = \eta$. □

DEFINITION 4.5. *Let $\psi \in \Psi$. Then we set*

$$\mathrm{TV}_\psi(u) := \int_\Omega \psi(\|\nabla u(x)\|)\, dx + \psi^\infty |D^s u|(\Omega), \quad (u \in \mathrm{BV}(\Omega)).$$

Remark 4.2. It can be shown that

$$\mathrm{TV}_\psi(u) = \sup\left\{ \int_\Omega u(x) \operatorname{div} \varphi(x)\, dx - \int_\Omega \psi^*(\|\varphi(x)\|)\, dx \;\middle|\; \varphi \in C_c^\infty(\Omega; \mathbb{R}^m)\right\}. \tag{4.6}$$

We now claim that TV_ψ satisfies the requirements of our main results, Theorem 3.7 and Theorem 3.8.

PROPOSITION 4.6. *Let $\Omega \subset \mathbb{R}^m$ be a bounded domain with Lipschitz boundary and $\psi \in \Psi$. Then TV_ψ is an admissible (separably) double-Lipschitz comparable regularisation functional on $L^1(\Omega)$. Moreover, $\|u\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)}$ for solutions u to* (P).

For the proof, we require the estimate from the following lemma.

LEMMA 4.7. *Let $\psi \in \Psi$ and $A, B : \mathbb{R}^m \to \mathbb{R}^m$ be linear transformations, and $c, d > 0$. Denote*

$$\begin{aligned}
\widetilde{G}_{A,B} &:= \sup_{\|v\|=1} \|Av\| + \|Bv\| - 2\|v\|,\\
\widetilde{J}_{c,d} &:= |c + d - 2|, \quad \textit{and}\\
\widetilde{D}_A &:= \|A - I\|.
\end{aligned}$$

If $\widetilde{J}_{c,d}, \widetilde{D}_A, \widetilde{D}_B \leq M$ for some constant $M \in (0, 1)$, then there exists a constant $C = C(M, K_\psi, \psi^\infty)$ such that

$$c\psi(\|Av\|) + d\psi(\|Bv\|) - 2\psi(\|v\|) \leq C(\widetilde{G}_{cA,dB} + \widetilde{J}_{c,d} + \widetilde{D}_A^2 + \widetilde{D}_B^2)\|v\|, \tag{4.7}$$

for all $v \in \mathbb{R}^m$.

Proof. Let us first of all observe directly from the definition of the subdifferential, and ψ being increasing that

$$0 < \sup_t \|\partial\psi(t)\| \leq \psi^\infty, \tag{4.8}$$

From convexity also

$$\psi(s) \le \psi(0) + t\frac{\psi(t/\lambda)}{t/\lambda}, \quad (\lambda \in (0,1]).$$

Letting $\lambda \searrow 0$, and recalling $\psi(0) = 0$, we deduce

$$\psi(s) \le \psi^\infty s, \quad (s \ge 0).$$

Let us define

$$L := c\psi(\|Av\|) + d\psi(\|Bv\|) - 2\psi(\|v\|).$$

We want to bound L. Using the definition of the subdifferential, we have

$$L \le c\partial\psi(\|Av\|)(\|Av\| - \|v\|) + d\partial\psi(\|Bv\|)(\|Bv\| - \|v\|) + (c+d-2)\psi(\|v\|). \quad (4.9)$$

If $\|Av\|, \|Bv\| \ge \|v\|$, we immediately deduce using (4.8) that

$$\begin{aligned} L &\le \psi^\infty \left(c\|Av\| + d\|Bv\| - 2\|v\|\right) + (c+d-2)\big(\psi(\|v\|) - \psi^\infty\|v\|\big) \\ &\le \psi^\infty\big(\widetilde{G}_{cA,dB}\|v\| + \widetilde{J}_{c,d}\|v\|\big). \end{aligned}$$

This is what we need.

If $\|Av\|, \|Bv\| \le \|v\|$, we deduce $L \le \psi^\infty \widetilde{J}_{c,d}\|v\|$. Again the claim holds.

It remains to consider the case $\|Av\| \ge \|v\| > \|Bv\|$, the case with A and B exchanged being analogous. We now use (4.5) as follows. We pick $z_B \in \partial\psi(\|Bv\|)$ and $z_A \in \partial\psi(\|Av\|)$, and define

$$\widehat{\chi} := 1 - \chi_{[K_\psi,\infty)}(\|Bv\|) = 1 - \chi_{[K_\psi,\infty)}(\|Av\|)\chi_{[K_\psi,\infty)}(\|Bv\|).$$

Then

$$\begin{aligned} z_B(\|Bv\| - \|v\|) &= \big(z_B - z_A\big)\big(\|Bv\| - \|v\|\big) + z_A\big(\|Bv\| - \|v\|\big) \\ &= \big(z_B - z_A\big)\big(\|Bv\| - \|Av\|\big) \\ &\quad + \big(z_B - z_A\big)\big(\|Av\| - \|v\|\big) + z_A\big(\|Bv\| - \|v\|\big) \\ &\le \widehat{\chi}C_\psi\big(\|Bv\| - \|Av\|\big)^2 + z_A\big(\|Bv\| - \|v\|\big). \end{aligned}$$

In the final step we have used (4.5), $\|Av\| - \|v\| \ge 0$, and the fact that $z_B - z_A \le 0$ which follows from the monotonicity of $\partial\psi$. Thus, continuing from (4.9), we calculate

$$\begin{aligned} L &\le cz_A(\|Av\| - \|v\|) + dz_B(\|Bv\| - \|v\|) + (c+d-2)\psi(\|v\|) \\ &\le cz_A(\|Av\| - \|v\|) + dz_A(\|Bv\| - \|v\|) + (c+d-2)\psi(\|v\|) \\ &\quad + d\widehat{\chi}C_\psi\big(\|Bv\| - \|Av\|\big)^2 \\ &\le z_A(\|cAv\| + \|dBv\| - 2\|v\|) + (c+d-2)\big(\psi(\|v\|) - z_A\|v\|\big) \\ &\quad + 2d\widehat{\chi}C_\psi\big((\|Av\| - \|v\|)^2 + (\|Bv\| - \|v\|)^2\big) \\ &\le \psi^\infty\big(\widetilde{G}_{cA,dB} + \widetilde{J}_{c,d}\big)\|v\| + 2(2+M)\widehat{\chi}C_\psi(\widetilde{D}_A^2 + \widetilde{D}_B^2)\|v\|^2. \end{aligned}$$

In the final step, we have used $d \le 2 + M$, which follows from the bound $\widetilde{J}_{c,d} \le M$ and $c > 0$. Finally, we observe from

$$\|Bv\| \ge \|v\| - \|Bv - v\| \ge (1-M)\|v\|$$

that

$$\widehat{\chi}\|v\| \le \frac{1}{1-M}\widehat{\chi}\|Bv\| \le \frac{K_\psi}{1-M}.$$

This proves the claim. □

Remark 4.3. We only needed (4.5) and the bound M on $\widetilde{D}_A$ and $\widetilde{D}_B$ for the estimate $\|v\|^2 \le C\|v\|$ below K_ψ. The bound M on $\widetilde{J}_{c,d}$ was used to get rid of d in the final steps.

Proof of Proposition 4.6. Weak* lower semicontinuity is immediate from the formulation (4.6). To see (3.1), we observe for some constants $c, r > 0$ that

$$\psi^*(t) \le c + \delta_{[0,r)}(t).$$

Indeed, for $t, s > K_\psi$, (4.5) implies $\partial\psi(s) = \partial\psi(t) = \{r\}$ for some constant $r > 0$. Approximating

$$\psi^*(t) = \sup_{s\ge 0}(st - \psi(s)) \le \sup_{s\ge 0}(st - r(s-2K_\psi) - \psi(2K_\psi))$$

and using $\psi(2K_\psi) < \infty$ now gives the above bound. Let us then pick $\varphi \in C_c^\infty(\Omega;\mathbb{R}^m)$ with $\|\phi\|_{2,L^\infty(\Omega)} \le r$. We estimate

$$\int_\Omega u \operatorname{div}\varphi\,dx - \int_\Omega \psi^*(\|\varphi(x)\|)\,dx \ge \int_\Omega u \operatorname{div}\varphi\,dx - c\mathcal{L}^m(\Omega).$$

It follows

$$r^{-1}\mathrm{TV}_\psi(\Omega) \ge |Du|(\Omega) - c\mathcal{L}^m(\Omega).$$

This shows (3.1) and that the assumptions of Definition 3.1 hold.

The proof of the assumptions of Definition 3.3 is analogous to Proposition 4.1. Let $\gamma \in \mathcal{F}(\Omega, U)$ for some open $U \subset \Omega$. The singular part $|D^s\gamma_\# u|(\Omega)$ is unaltered in (4.1). In place of the absolutely continuous part $|D^a\gamma_\# u|(\Omega)$, we have $\int_\Omega \psi(\|\nabla\gamma_\# u\|)\,dx$. Using the area formula, we calculate

$$\int_\Omega \psi(\|\nabla\gamma_\# u(x)\|)\,dx = \int_U \psi(\|\nabla\gamma^{-1}(\gamma(x))\nabla u(x)\|)\mathcal{J}_m\gamma(x)\,dx.$$

Let now $\overline{\gamma}, \underline{\gamma} \in \mathcal{F}(\Omega, U)$ with $T_{\overline{\gamma},\underline{\gamma}} < 1$. Summing the previous equation for $\gamma = \overline{\gamma}, \underline{\gamma}$, subtracting $2\int_\Omega \psi(\|\nabla u(x)\|)\,dx$ and using Lemma 4.7 with $A = \nabla\overline{\gamma}^{-1}(\gamma(x))$, $B = \nabla\underline{\gamma}^{-1}(\gamma(x))$, $c = \mathcal{J}_m\overline{\gamma}(x)$, $d = \mathcal{J}_m\underline{\gamma}(x)$, and $v = \nabla u(x)$ yields

$$\begin{aligned}\int_\Omega \psi(\|\nabla\overline{\gamma}_\# u\|)\,dx + \int_\Omega \psi(\|\nabla\underline{\gamma}_\# u\|)\,dx - 2\int_\Omega \psi(\|\nabla u(x)\|)\,dx \\ \le C(G_{\overline{\gamma},\underline{\gamma}} + \bar{J}_{\overline{\gamma},\underline{\gamma}} + D^2_{\overline{\gamma}} + D^2_{\underline{\gamma}})\int_U \|\nabla u(x)\|\,dx.\end{aligned}$$

The various elements in the sum on the right hand side are defined in (3.2). With this estimate at hand, the rest follows exactly as in Proposition 4.1.

Finally, to see, $\|u\|_{L^\infty(\Omega)} \le \|f\|_{L^\infty(\Omega)}$ for solutions u to (P), we follow the corresponding proof for TV. Namely, if we set $L := \|f\|_{L^\infty(\Omega)}$, and replace u by

$$\bar{u}(x) := \max\{-L, \min\{u(x), L\}\},$$

then it follows from the chain rule in BV [3, Theorem 3.96] that

$$|D\bar{u}|(A) \leq |Du|(A) \quad \text{for any Borel set } A.$$

From this it is immediate that

$$\mathrm{TV}_\psi(\bar{u}) \leq \mathrm{TV}_\psi(u),$$

Moreover, if $\bar{u} \neq u$, i.e., $\|\bar{u}\|_{L^\infty(\Omega)} < \|u\|_{L^\infty(\Omega)}$, we deduce

$$\int_\Omega \phi(\bar{u}(x) - f(x))\, dx < \int_\Omega \phi(u(x) - f(x))\, dx.$$

This provides a contradiction. Thus we must have $\|u\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)}$. □

4.3. Remarks on ill-posed non-convex regularisers. We now provide a few remarks on using the technique on popular models involving non-convex energies ψ. Problems of the form (P) with regularisation functionals TV_ψ employing the energies constructed below do not, however, in general have solutions in $\mathrm{BV}(\Omega)$. Some remedies exist [25] which make these models still worth considering.

Remark 4.4. (Non-convex total variation) Regularisation functionals TV_ψ based on concave energies ψ have recently received increased attention, for the better modelling of real image gradient statistics [26, 23, 24, 31, 25]. Let us define

$$\psi^0 := \lim_{t \searrow 0} \psi(t)/t,$$

and suppose $\psi : [0,\infty) \to [0,\infty)$ is concave, increasing, $\psi(0) = 0$ and $\psi^0, \psi^\infty \in (0,\infty)$. The upper bound on ψ^0 forbids the typical energies $\psi(t) = t^q$ for $q \in (0,1)$, but allows slightly modified ones, linearised for small values. The above proof can now be extended to show that TV_ψ is still (separably) double-Lipschitz comparable. Indeed, Lemma 4.7 is trivial to prove in this case. Assuming for simplicity that $c + d = 2$, by concavity

$$\begin{aligned} c\psi(\|Av\|) + d\psi(\|Bv\|) - 2\psi(\|v\|) &\leq c\partial\psi(\|v\|)(\|Av\| - \|v\|) + d\partial\psi(\|v\|)(\|Bv\| - \|v\|) \\ &\leq \psi^0 \max\{0, \|cAv\| + \|dBv\| - 2\|v\|\} \\ &\leq \psi^0 \widetilde{G}_{A,B}\|v\|. \end{aligned}$$

The assumption $c + d = 2$ is also not difficult to remove with the help of $\widetilde{J}_{c,d}$, and is satisfied by the Lipschitz transformation we construct in §5.

Unfortunately, the TV_ψ regularisation functional constructed with concave ψ is not admissible. It lacks weak* lower semicontinuity. Using area strict convergence [36] and additional multiscale regularisation, first introduced in [40] for discontinuous optical flow, problems involving such regularisation can however be made well-posed [25]. It is outside the scope of the present paper to prove that the multiscale regularisation term η is separably double-Lipschitz comparable. Nevertheless, assuming we had a solution $u \in \mathrm{BV}(\Omega)$ to the basic model, or a remedied model still satisfying the double-Lipschitz comparability condition, then our technique would show $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$.

Remark 4.5. (Perona-Malik) The Perona-Malik anisotropic diffusion [35] may also be written in the variational form (P) with $\phi(t) = t^2$ and $R(u) = \mathrm{TV}_\psi(u)$ for $\psi(t) = \log(1 + t^2)$. Observe that this ψ is not even concave, unlike the models of the previous remark. The model suffers from exactly the same ill-posedness problems; for a review on approaches to make the problem well-posed, we refer to [21]. Nevertheless, our approach can be adapted to study it, assuming we have a solution $u \in \mathrm{BV}(\Omega)$ to the problem. The existence is however generally not guaranteed. Indeed, with the notation of Lemma 4.7, by properties of the logarithm, we have

$$c\psi(\|Av\|) = c\log(1+\|Av\|^2) = c\log(c + c\|Av\|^2) - c\log c. \tag{4.10}$$

By the concavity of the logarithm, we have

$$\begin{aligned} c\log(c + c\|Av\|^2) - c\log(1+\|v\|^2) &\le \frac{c}{1+\|v\|^2}(c\|Av\|^2 - \|v\|^2 + c - 1) \\ &= \frac{1}{1+\|v\|^2}\big(\|cAv\|^2 - \|v\|^2 + (c-1)(\|v\|^2+1)\big). \end{aligned}$$

If $c + d = 2$, summing with the corresponding estimate for B and d, we have

$$c\log(c + c\|Av\|^2) + d\log(d + d\|Bv\|^2) - 2\log(1+\|v\|^2) \le \frac{1}{1+\|v\|^2} G'_{cA,dB}\|v\|^2,$$

where

$$G'_{cA,dB} = \|c^2 A^* A + d^2 B^* B - 2I\|.$$

With $c + d = 2$, we also have

$$-c\log c - d\log d \le 0.$$

Therefore, using $1 + \|v\|^2 \ge \|v\|$ and referring back to (4.10), we obtain

$$c\psi(\|Av\|) + d\psi(\|Bv\|) - 2\psi(\|v\|) \le G'_{cA,dB}\|v\|.$$

This is not exactly the estimate of Lemma 4.7, but in §5, we will in fact estimate $\widetilde{G}_{cA,dB}$ through $G'_{cA,dB}$; see Lemma 5.1. Therefore we may follow this reasoning with the argument of Proposition 4.6. We also assumed $c + d = 2$, but it would not be difficult to remove this assumption with the help of $\widetilde{J}_{c,d}$. The specific Lipschitz transformations that we construct next in §5 however do satisfy $c + d = 2$, that is, $\mathcal{J}_m\overline{\gamma}(x) + \mathcal{J}_m\underline{\gamma}(x) = 2$.

5. Lipschitz shift transformations. We now introduce the "shift" class of Lipschitz transformations that we will use to push forward a purported solution u to (P) that does not satisfy $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$. The idea is to take a Lipschitz graph Γ on J_u containing a violating point x_0, and then to move the jump forward or backward by an amount ρ, in order to construct a better candidate solution. It is interesting to relate the specific constructions here to the general framework for designing transformations satisfying a PDE on the Jacobian determinant in [15]. Before the construction, we provide a simple general lemma stating one of the most crucial parts of our technique. So far, we have not made significant use of the fact that our finite-dimensional norm on $\mathbb{R}^m$ is the Euclidean norm, but here it is essential.

5.1. A crucial estimate. We will use the following simple lemma to simplify the estimation of the double-Lipschitz comparison factor $G_{\overline{\gamma},\underline{\gamma}}$.

LEMMA 5.1. *Let $\overline{\gamma}, \underline{\gamma} \in \mathcal{F}(\Omega)$. For $G_{\overline{\gamma},\underline{\gamma}}$ defined in (3.2a), and A_γ defined in (3.2c), we have*

$$G_{\overline{\gamma},\underline{\gamma}} \le \sup_{x\in\Omega} \frac{1}{2}\|[A_{\overline{\gamma}}(x)]^* A_{\overline{\gamma}}(x) + [A_{\underline{\gamma}}(x)]^* A_{\underline{\gamma}}(x) - 2I\|.$$

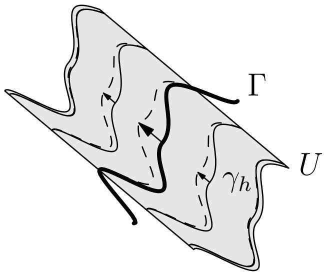

FIGURE 5.1. *Illustration of the transformation γ_h constructed in Lemma 5.2. The solid lines are transformed into the dotted lines, with larger movement deep in the interior of U (gray region), and none on the boundary.*

Proof. By the concavity of the square root, for any $0 \neq v \in \mathbb{R}^m$, we have

$$\|A_\gamma(x)v\| - \|v\| \leq \frac{1}{2\|v\|}\left(\|A_\gamma(x)v\|^2 - \|v\|^2\right) = \frac{1}{2\|v\|}\langle v, ([A_\gamma(x)]^* A_\gamma(x) - I)v\rangle.$$

Therefore

$$\begin{aligned}\|A_{\overline{\gamma}}(x)v\| + \|A_{\underline{\gamma}}(x)v\| - 2\|v\| &\leq \frac{1}{2\|v\|}\langle v, ([A_{\overline{\gamma}}(x)]^* A_{\overline{\gamma}}(x) + [A_{\underline{\gamma}}(x)]^* A_{\underline{\gamma}}(x) - 2I)v\rangle \\ &\leq \frac{1}{2}\|[A_{\overline{\gamma}}(x)]^* A_{\overline{\gamma}}(x) + [A_{\underline{\gamma}}(x)]^* A_{\underline{\gamma}}(x) - 2I\|\|v\|.\end{aligned}$$

The lemma is proved. □

5.2. The construction. For brevity, we now define $\mathfrak{v}_x := P_{z_\Gamma}^{\perp} x$ and $\mathfrak{t}_x := \langle z_\Gamma, x\rangle$ to be the components of x on V_Γ and along z_Γ, for a Lipschitz graph Γ. The transformation γ_h constructed in the next lemma is illustrated in Figure 5.1.

LEMMA 5.2. *Let $\Gamma \subset \Omega$ be a Lipschitz $(m-1)$-graph and $s > 0$ be such that*

$$U := \Gamma + (-s, s)z_\Gamma \subset \Omega.$$

Suppose that $h \in W_0^{1,\infty}(V_\Gamma)$ satisfies $-s/3 \leq h \leq s/3$. Define the transformation $\gamma_h : U \to U$ by

$$\gamma_h(x) := \mathfrak{v}_x + \widetilde{\gamma}_{s,h}(\mathfrak{t}_x; \mathfrak{v}_x) z_\Gamma \tag{5.1}$$

where

$$\widetilde{\gamma}_{s,h}(t; v) := \begin{cases} t + \frac{s - f_\Gamma(v) + t}{s} h(v), & f_\Gamma(v) - s < t < f_\Gamma(v), \\ t + \frac{s + f_\Gamma(v) - t}{s} h(v), & f_\Gamma(v) \leq t < f_\Gamma(v) + s. \end{cases}$$

Then γ_h is 1-to-1 and Lipschitz and

$$\mathcal{J}_m \gamma_h(x) = |\widetilde{\gamma}'_{s,h}(\mathfrak{t}_x; \mathfrak{v}_x)| = \begin{cases} 1 + \frac{h(\mathfrak{v}_x)}{s}, & f_\Gamma(\mathfrak{v}_x) - s < \mathfrak{t}_x < f_\Gamma(\mathfrak{v}_x), \\ 1 - \frac{h(\mathfrak{v}_x)}{s}, & f_\Gamma(\mathfrak{v}_x) \leq \mathfrak{t}_x < f_\Gamma(\mathfrak{v}_x) + s. \end{cases}$$

There also exists a constant $C = C(h, f_\Gamma) > 0$ such that for $\rho \in (-1, 1)$, we have for the various double-Lipschitz comparison constants defined in (3.2) the estimates

$$\begin{aligned} &T_{\gamma_{\rho h}, \gamma_{-\rho h}} \leq C\rho^2, && G_{\gamma_{\rho h}, \gamma_{-\rho h}} \leq C\rho^2, && \bar{J}_{\gamma_{\rho h}, \gamma_{-\rho h}} = 0, && D_{\gamma_{\rho h}} \leq C|\rho|, \\ &T_{\gamma_{\rho h}, \iota} \leq C|\rho|, && G_{\gamma_{\rho h}, \iota} \leq C|\rho|, && \bar{J}_{\gamma_{\rho h}, \iota} \leq C|\rho|, && \\ &T_{\gamma_{\rho h}^{-1}, \iota} \leq C|\rho|, && G_{\gamma_{\rho h}^{-1}, \iota} \leq C|\rho|, && \bar{J}_{\gamma_{\rho h}^{-1}, \iota} \leq C|\rho|, && D_{\gamma_{\rho h}^{-1}} \leq C|\rho|. \end{aligned} \tag{5.2}$$

Remark 5.1. It does not hold that $\gamma_h^{-1} = \gamma_{-h}$, although it would be possible to construct such a transformation satisfying the same essential properties. We could then simply plug it into our proofs instead of the above one. We do not however do this, since the bounds (5.2) are in that case a bit more work to prove, and having such a transformation would only help very little with Lemma B.3.

Proof. After rotation and translation, if necessary, we may assume w.log that $z_\Gamma = (0,\ldots,0,1)$, so that $x = (v,t) := (\mathfrak{v}_x, \mathfrak{t}_x)$. Then

$$\widetilde{\gamma}'_{s,h}(t;v) = d_h := \begin{cases} 1 + \frac{h(v)}{s}, & f_\Gamma(v) - s < t < f_\Gamma(v), \\ 1 - \frac{h(v)}{s}, & f_\Gamma(v) \le t < f_\Gamma(v) + s. \end{cases}$$

and

$$\nabla_v \widetilde{\gamma}_{s,h}(t;v) = c_h := \begin{cases} \frac{s - f_\Gamma(v) + t}{s} \nabla h(v) - \frac{\nabla f_\Gamma(v)}{s} h(v), & f_\Gamma(v) - s < t < f_\Gamma(v), \\ \frac{s + f_\Gamma(v) - t}{s} \nabla h(v) + \frac{\nabla f_\Gamma(v)}{s} h(v), & f_\Gamma(v) \le t < f_\Gamma(v) + s. \end{cases}$$

Observe that $|\widetilde{\gamma}'_{s,h}(t;v)| = \widetilde{\gamma}'_{s,h}(t;v)$ due to the bounds $-s/3 \le h \le s/3$.

To calculate the Lipschitz factor and the Jacobian determinant

$$\mathcal{J}_m \gamma_h(x) = \sqrt{\det(\nabla\gamma_h(x)[\nabla\gamma_h(x)]^T)},$$

we study the eigenvalues $\lambda_1, \ldots, \lambda_m$ of

$$\nabla\gamma_h(x)[\nabla\gamma_h(x)]^T = \begin{pmatrix} I & c_h \\ 0 & d_h \end{pmatrix} \begin{pmatrix} I & 0 \\ c_h^T & d_h \end{pmatrix} = \begin{pmatrix} I + c_h \otimes c_h & d_h c_h \\ d_h c_h^T & d_h^2 \end{pmatrix}.$$

Easily we see that $\lambda_3 = \cdots = \lambda_m = 1$, with the corresponding eigenvector orthogonal to c_h. For the two remaining, important, eigenvalues, setting $y = (c_h, \alpha)$ for the unknown eigenvector, we obtain the system of equations

$$1 + \|c_h\|^2 + \alpha d_h = \lambda \quad \text{and} \quad d_h \|c_h\|^2 + \alpha d_h^2 = \lambda\alpha.$$

Solving this system of equations, we obtain the solutions

$$\lambda_1(x), \lambda_2(x) = \frac{1 + d_h^2 + \|c_h\|^2 \pm \sqrt{(1 + d_h^2 + \|c_h\|^2)^2 - 4d_h^2}}{2}.$$

This gives as claimed

$$\mathcal{J}_m \gamma_h(x) = \sqrt{\prod_{i=1}^m \lambda_i} = \sqrt{\lambda_1(x)\lambda_2(x)} = |d_h| = \widetilde{\gamma}'_{s,h}(t;v).$$

It remains to consider the bounds (5.2). Letting $\overline{\gamma} := \gamma_{\rho h}$ and $\underline{\gamma} := \gamma_{-\rho h}$, and recalling that

$$A_\gamma(x) := \nabla\gamma^{-1}(\gamma(x)) \mathcal{J}_m \gamma(x),$$

by Lemma 5.1 we have

$$G_{\overline{\gamma},\underline{\gamma}} \le \sup_{x \in \Omega} \frac{1}{2} \|[A_{\overline{\gamma}}(x)]^* A_{\overline{\gamma}}(x) + [A_{\underline{\gamma}}(x)]^* A_{\underline{\gamma}}(x) - 2I\|.$$

We want to estimate this further. Recalling the proof of Lemma 4.2, $\nabla\gamma_h^{-1}(\gamma_h(x)) = [\nabla\gamma_h(x)]^{-1}$ holds (a.e.). With

$$\nabla\gamma_h(x) = \begin{pmatrix} I & 0 \\ c_h^T & d_h \end{pmatrix},$$

it can easily be verified that

$$[\nabla\gamma_h(x)]^{-1} = \begin{pmatrix} I & 0 \\ -c_h^T d_h^{-1} & d_h^{-1} \end{pmatrix}. \tag{5.3}$$

It follows that

$$A_{\gamma_h}(x) = \begin{pmatrix} d_h I & 0 \\ -c_h^T & 1 \end{pmatrix},$$

and

$$[A_{\gamma_h}(x)]^* A_{\gamma_h}(x) = \begin{pmatrix} d_h^2 I + c_h^T c_h & -c_h \\ -c_h^T & 1 \end{pmatrix}.$$

We observe that

$$c_{\pm\rho h} = \pm\rho c_h, \quad \text{and} \tag{5.4}$$

$$d_{\pm\rho h}^2 = 1 \pm 2\rho d_h + \rho^2 d_h^2, \tag{5.5}$$

Thus $d_{\rho h}^2 + d_{-\rho h}^2 - 2 = 2\rho^2 d_h^2$, and we deduce

$$[A_{\overline{\gamma}}(x)]^* A_{\overline{\gamma}}(x) + [A_{\underline{\gamma}}(x)]^* A_{\underline{\gamma}}(x) - 2I = 2\rho^2 \begin{pmatrix} d_h^2 I + c_h^T c_h & 0 \\ 0 & 0 \end{pmatrix}.$$

It follows

$$G_{\overline{\gamma},\underline{\gamma}} \le C\rho^2.$$

Moreover, (5.5) also yields

$$\bar{J}_{\gamma_{\rho h},\gamma_{-\rho h}} = d_{\rho h} + d_{-\rho h} - 2 = 0.$$

In order to estimate

$$D_{\gamma_{\rho h}} = \sup_{x\in\Omega} \|\nabla\gamma_{\rho h}^{-1}\gamma_{\rho h}(x) - I\|,$$

we calculate using (5.3) that

$$[\nabla\gamma_{\rho h}(x)]^{-1} - I = d_{\rho h}^{-1} \begin{pmatrix} 0 & 0 \\ -c_{\rho h}^T & 1 - d_{\rho h}. \end{pmatrix}$$

By the bounds on r, f_Γ and h, we have $|d_{\rho h}| \ge 2/3$. Therefore, it follows from (5.4), (5.5) that

$$D_\gamma = \sup_{x\in\Omega} \|\nabla\gamma_{\rho h}^{-1}(\gamma_{\rho h}(x)) - I\| = \sup_{x\in\Omega} \|[\nabla\gamma_{\rho h}(x)]^{-1} - I\| \le C\rho, \quad (0 < \rho < 1).$$

This proves all the bounds in (5.2) involving both $\gamma_{\rho h}$ and $\gamma_{\rho h}$. The bounds involving ι and $\gamma_{\rho h}^{-1}$ are proved analogously. For the $\gamma_{\rho h}^{-1}$ bounds we use the fact at $y = \gamma_h(x)$ we have

$$\mathcal{J}_m \gamma_h^{-1}(y) = 1/\mathcal{J}_m \gamma_h(x) = 1/d_h,$$

as well as $\nabla \gamma_h(\gamma_h^{-1}(y)) = \nabla \gamma_h(x)$, so that

$$A_{\gamma_h^{-1}} = \begin{pmatrix} d_h^{-1} I & 0 \\ -c_h^T d_h^{-1} & 1 \end{pmatrix}.$$

We skip the elementary details. □

We will also require an estimate of the following type.

LEMMA 5.3. *Let the Lipschitz transformation $\gamma_h : \Omega \to \Omega$ have the form (5.1), and be identity outside $U \subset \Omega$. Let $u \in \mathrm{BV}(\Omega)$. Define*

$$M_\gamma := \sup_{x \in U} \|\gamma_h(x) - x\|.$$

Then $M_\gamma = \|h(v)\|_{L^\infty(V_\Gamma)}$ and

$$\int_U |u(\gamma_h(x)) - u(x)| \, dx \le M_{\gamma_h} |Du|(U).$$

Proof. Let us again, without loss of generality after rotation and translation, if necessary, that $z_\Gamma = (0, \ldots, 0, 1)$, so that $x = (v, t) := (\mathfrak{v}_x, \mathfrak{t}_x)$. Then

$$\gamma_h(v, t) = (v, \widetilde{\gamma}_v(t)),$$

and the slice

$$u_v(t) := u(v, t), \quad (-s < t < s; v \in V_\Gamma)$$

satisfies $u_v \in \mathrm{BV}(-s, s)$ for $\mathcal{L}^{m-1}$-a.e. $v \in V_\Gamma$. Thus

$$|u_v(\widetilde{\gamma}(t; v)) - u_v(t)| \le |Du_v|([t, \widetilde{\gamma}(t; v)]) = \int_{-s}^{s} \chi_{[t, \widetilde{\gamma}(t;v)]}(\tau) \, d|Du_v|(\tau).$$

Here we use the convention $[a, b] := [b, a]$ if $b < a$. Observe that $[t, \widetilde{\gamma}(t; v)] \subset [t, t + M_{\gamma_h}]$ or $[t, \widetilde{\gamma}(t; v)] \subset [t, t - M_{\gamma_h}]$. Using Fubini's theorem and basic properties of one-dimensional slices in BV [3], we may thus estimate

$$\begin{aligned}
\int_U |u(\gamma_h(x)) - u(x)| \, dx &= \int_{V_\Gamma} \int_{-s}^{s} |u_v(\gamma_v(t)) - u_v(t)| \, dt \, d\mathcal{H}^{m-1}(v) \\
&\le \int_{V_\Gamma} \int_{-s}^{s} \int_{-s}^{s} \chi_{[t, \widetilde{\gamma}(t;v)]}(\tau) \, d|Du_v|(\tau) \, dt \, d\mathcal{H}^{m-1}(v) \\
&= \int_{V_\Gamma} \int_{-s}^{s} \int_{-s}^{s} \chi_{[t, \widetilde{\gamma}(t;v)]}(\tau) \, dt \, d|Du_v|(\tau) \, d\mathcal{H}^{m-1}(v) \\
&\le \int_{V_\Gamma} \int_{-s}^{s} M_{\gamma_h} \, d|Du_v|(\tau) \, d\mathcal{H}^{m-1}(v) \\
&= M_{\gamma_h} |Du|(U).
\end{aligned}$$

Finally

$$\|\gamma_h(x) - x\| = |h(v)| \cdot \begin{cases} \frac{s - f_\Gamma(v) + t}{s}, & f_\Gamma(v) - s < t < f_\Gamma(v), \\ \frac{s + f_\Gamma(v) - t}{s}, & f_\Gamma(v) \le t < f_\Gamma(v) + s. \end{cases}$$

Clearly this gives $M_{\gamma_h} = \|h(v)\|_{L^\infty(V_\Gamma)}$. □

In the following lemma, based on the construction of Lemma 5.2, we now construct our family of shift transformations parametrised by the size $r > 0$ of the neighbourhood of x_0 where the transformation is performed, and the magnitude $\rho > 0$ of the transformation.

Lemma 5.4. *Let $\Omega \subset \mathbb{R}^m$, and $\Gamma \subset \Omega$ be a Lipschitz $(m-1)$-graph. Pick $x_0 \in \Gamma$ and $0 \neq \bar{h} \in W_0^{1,\infty}(z_\Gamma^\perp \cap B(0,1))$ satisfying $0 \le \bar{h} \le 1$. Set*

$$h_r(v) := r\bar{h}((v - \mathfrak{v}_{x_0})/r), \quad (r > 0).$$

Then there exists $r_0 > 0$ such that each of the maps

$$\gamma_{\rho,r}(x) := \mathfrak{v}_x + \widetilde{\gamma}_{3r,\rho h_r}(\mathfrak{t}_x; \mathfrak{v}_x) z_\Gamma, \quad (-1 < \rho < 1,\, 0 < r < r_0),$$

satisfies $\gamma_{\rho,r} \in \mathcal{F}(\Omega; U_r)$ for

$$U_r := x_0 + z_\Gamma^\perp \cap B(0,r) + (3 + \operatorname{lip} f_\Gamma)(-r,r) z_\Gamma.$$

Moreover, there exists a constant $C > 0$ such that the transformations $\gamma_{\rho,r}$, $(0 < r < r_0,\ -1 < \rho < 1)$, satisfy

$$T_{\gamma_{\rho,r},\gamma_{-\rho,r}} \le C\rho^2, \quad T_{\gamma_{\rho,r},\iota} \le C|\rho|,$$
$$M_{\gamma_{\rho,r}} = |\rho| r, \quad T_{\gamma_{\rho,r}^{-1},\iota} \le C|\rho|$$

for the various double-Lipschitz comparison constants defined in (3.2).

Remark 5.2. We can, for example, set $\bar{h}(v) := \max\{0, 1 - \|v\|\}$.

Proof. We choose $r_0 > 0$ small enough that $U_r \subset \Omega$. Clearly $\gamma_{\rho,r}$ and $\gamma_{\rho,r}^{-1}$ satisfy the Lusin N-property, and are Lipschitz mappings. Since $0 \le \bar{h} \le 1$, evidently also $M_\gamma = |\rho| r$, and γ reduces to the identity outside U_r. The remaining claims follow by applying Lemma 5.2 on

$$\Gamma' := g_\Gamma(V_\Gamma \cap B(\mathfrak{v}_{x_0}, r)) \subset \Gamma$$

with $h = h_r$ and $s = 3r$. □

6. Proof of the main result for $p > 1$. We now begin the proof of Theorem 3.7. Most of the proof consists of producing for the fidelity term a counterpart of the double-Lipschitz comparability condition of the regulariser. We do this in §6.2 after a technical lemma in §6.1. The fidelity estimates are based on the specific Lipschitz transformations constructed in the previous section. Averaging over two different Lipschitz transformations $\overline{\gamma} = \gamma_{\rho,r}$ and $\underline{\gamma} = \gamma_{-\rho,r}$ will provide an $O(\rho)$ decrease estimate for the fidelity. Importantly, in order to take advantage of the strict convexity of ϕ, we actually need as our improvement candidates convex combinations $\theta u + (1-\theta)\gamma_\# u$. After dealing with the fidelity function, we apply in §6.3 the double-Lipschitz comparability to get an $O(\rho^2)$ increase estimate on the regulariser – averaged over the two transformations. After these estimates, the proof of Theorem 3.7 will be immediate.

6.1. A technical lemma. We can only perform fidelity estimation at a base point x_0 outside the set Z_u we construct next. For this, given $u \in \mathrm{BV}(\Omega)$, we now fix a countably family of Lipschitz $(m-1)$-graphs $\{\Gamma_i\}_{i=1}^{\infty}$, such that

$$\mathcal{H}^{m-1}(J_u \setminus \bigcup_{i=1}^{\infty} \Gamma_i) = 0. \tag{6.1}$$

Such a family exists by the rectifiability of J_u. For $\mathcal{H}^{m-1}$-a.e. $x_0 \in J_u$, we may then find a Lipschitz graph $\Gamma = \Gamma^{x_0}$ with $x_0 \in \Gamma^{x_0} \Subset \Gamma_i$ for some $i = i(x_0)$, satisfying the following two properties. Firstly $V_{\Gamma^{x_0}} \supset B(P_{z_\Gamma}^{\perp} x_0, r(x_0))$ for some $r(x_0) > 0$. This can clearly be satisfied since we can by Kirzbraun's theorem assume $V_\Gamma = z_\Gamma^{\perp}$. Secondly we can take the traces of u from both sides of Γ to exist at x_0 and agree with $u^{\pm}(x_0)$. This can be seen from, e.g., the BV trace theorem [3, Theorem 3.77].

Lemma 6.1. *Let $u \in \mathrm{BV}(\Omega)$. Then there exists a Borel set Z_u with $\mathcal{H}^{m-1}(Z_u) = 0$ such that every $x \in J_u \setminus Z_u$ is a Lebesgue point of the one-sided traces $u^{\pm}$, and*

$$\Theta_{m-1}^*(|Du|\llcorner(\Gamma^x)^+; x) = 0, \text{ and } \Theta_{m-1}^*(|Du|\llcorner(\Gamma^x)^-; x) = 0. \tag{6.2}$$

Proof. For $\mathcal{H}^{m-1}$-a.e. $x \in \Gamma_i$, we have both

$$\Theta_{m-1}(|Du|\llcorner\Gamma_i; x) = |u^+(x) - u^-(x)|,$$

and

$$\Theta_{m-1}(|Du|; x) = |u^+(x) - u^-(x)|,$$

as follows from [3, Theorem 2.83 & Theorem 3.77] and, for the latter, a simple application of the former and the Besicovitch derivation theorem. Since

$$\Theta_{m-1}^*(|Du|\llcorner(\Gamma_i^+ \cup \Gamma_i^-); x) \le \Theta_{m-1}(|Du|; x) < \infty$$

we see using the definition of Θ_{m-1}^* that

$$\Theta_{m-1}^*(|Du|\llcorner(\Gamma_i^+ \cup \Gamma_i^-); x) + \Theta_{m-1}(|Du|\llcorner\Gamma_i; x) = \Theta_{m-1}(|Du|; x).$$

Therefore

$$\Theta_{m-1}^*(|Du|\llcorner(\Gamma_i^+ \cup \Gamma_i^-); x) = 0.$$

This gives (6.2) for $\mathcal{H}^{m-1}$-ae $x \in J_u$. Finally $\mathcal{H}^{m-1}$-a.e. $x \in J_u$ is a Lebesgue point of the traces $u^{\pm}$, so clearly the set Z_u satisfying the claims exists. □

6.2. The effect of the shift transformation on the fidelity. Recalling the definition of $\widetilde{J}_f$ in (2.3), we now fix $x_0 \in J_u \setminus (\widetilde{J}_f \cup S_f \cup Z_u)$. We also let $\Gamma = \Gamma^{x_0}$, observing from the construction above that the traces of u from both sides of Γ exist at x_0 and agree with $u^{\pm}(x_0)$. We also recall the construction of Lemma 5.4. We are given a base Lipschitz function $\bar{h} \in W_0^{1,\infty}(z_\Gamma^{\perp} \cap B(0,1))$ with $0 \le \bar{h} \le 1$. (Sometimes we only assume $-1 \le h \le 1$. This is explicitly stated.) We set

$$h_r(v) := r h((v - \mathfrak{v}_{x_0})/r),$$

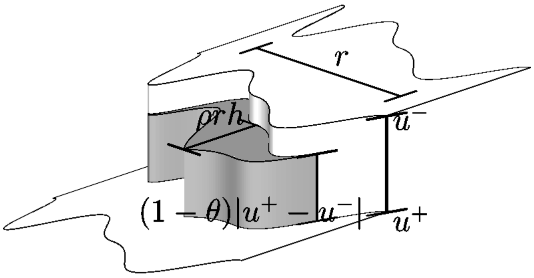

FIGURE 6.1. *Illustration of the function* $\bar{u}_{\rho,r}(x) := \theta u(x) + (1-\theta)\gamma_{\rho,r\#}u(x)$ *(gray) constructed in §6.2 in comparison to u (white).*

and define the Lipschitz transformations $\gamma_{\rho,r} : \Omega \to \Omega$ by

$$\gamma_{\rho,r}(x) := \mathfrak{v}_x + \widetilde{\gamma}_{3r,\rho h_r}(\mathfrak{t}_x; \mathfrak{v}_x) z_\Gamma, \quad (-1 < \rho < 1,\ 0 < r < r_0).$$

These transformations are the identity outside

$$U_r := x_0 + z_\Gamma^\perp \cap B(0,r) + (3 + \operatorname{lip} f_\Gamma)(-r,r) z_\Gamma,$$

which we split into the halves

$$U_r^\pm = U_r \cap \Gamma^\pm.$$

Here we choose $r > 0$ small enough that $U_r \subset \Omega$. We observe and put in our mind for later that

$$U_r \subset B(x_0, Cr) \quad \text{for} \quad C = \sqrt{1 + (3 + \operatorname{lip} f_\Gamma)^2}.$$

We pick arbitrary $\theta \in [0,1]$ as well as $r \in (0, r_0)$ and $\rho \in (0,1)$. With these we define

$$\bar{u}_{\rho,r}(x) := \theta u(x) + (1-\theta)\gamma_{\rho,r\#}u(x),$$

illustrated in Figure 6.1, as well as the piecewise constant functions

$$\begin{aligned} f_0 &:= \widetilde{f}(x_0), \\ u_0 &:= u^+(x_0)\chi_{U_r^+} + u^-(x_0)\chi_{U_r^-}, \quad \text{and} \\ \bar{u}_{\rho,r,0} &:= \theta u_0(x) + (1-\theta)\gamma_{\rho,r\#}u_0(x). \end{aligned}$$

We aim to prove that for suitable (θ, ρ, r), either $\bar{u}_{\rho,r}$ or $\bar{u}_{-\rho,r}$ is better than u. We do this by averaging estimates over the two piecewise constant functions. The interpolation parameter θ will in particular be necessary to take advantage of the p-increase (or strict convexity) of ϕ for $p > 1$.

First, we have to estimate the discrepancy between the piecewise constant functions and the original ones. Without assuming u bounded, we could get an $O(\epsilon r^m)$ estimate. We are not able to deal with this in the general analysis. We therefore have to assume u (locally) bounded, in order to improve this to $O(\epsilon \rho r^m)$.

LEMMA 6.2. *Suppose ϕ is p-increasing with $1 \le p < \infty$. Suppose either $u, f \in \mathrm{BV}(\Omega) \cap L^\infty_{\mathrm{loc}}(\Omega)$, or that $p = 1$ and $u, f \in \mathrm{BV}(\Omega)$. Let $x_0 \in J_u \setminus (\widetilde{J}_f \cup S_f \cup Z_u)$. Given*

$\epsilon > 0$, there exists $r_0 > 0$, independent of ρ and valid for every $\bar{h} \in W_0^{1,\infty}(z_\Gamma^\perp \cap B(0,1))$ with $-1 \le \bar{h} \le 1$, such that whenever $0 < r < r_0$ and $-1 < \rho < 1$, then

$$\begin{aligned}\int_\Omega \phi(\bar{u}_{\rho,r}(x) - f(x))\,dx - \int_\Omega \phi(u(x) - f(x))\,dx \\ \le \int_{U_r} \phi(\bar{u}_{\rho,r,0}(x) - f_0(x))\,dx - \int_{U_r} \phi(u_0(x) - f_0(x))\,dx + \epsilon|\rho| r^m.\end{aligned} \tag{6.3}$$

Proof. We may without loss of generality assume $\rho > 0$, since the case $\rho = 0$ is trivial, and the case $\rho < 0$ can be handled by negating $\bar{h}$. We begin the proof by choosing $r > 0$ small enough that $U_r \subset \Omega$. Further, if $p > 1$, we pick $r > 0$ small enough that u and f are essentially bounded within U_r. Since $\gamma_{\rho,r}$ is the identity outside U_r, it suffices to perform estimation with U_r. We do this with distinct arguments within the $O(\rho r^m)$ sets $W_r^+ := U_r^+ \cap \gamma(U_r^-)$ and $W_r^- := U_r^- \cap \gamma(U_r^+)$, and within the $O(r^m)$ sets $\widetilde{W}_r^+ := U_r^+ \cap \gamma(U_r^+)$ and $\widetilde{W}_r^- := U_r^- \cap \gamma(U_r^-)$. We begin with the latter, observing that $\bar{u}_{\rho,r,0}(x) = u_0(x)$ for $x \in \widetilde{W}_r^\pm$. Thus

$$\int_{\widetilde{W}^\pm} \phi(\bar{u}_{\rho,r,0}(x) - f_0(x)) - \phi(u_0(x) - f_0(x))\,dx = 0. \tag{6.4}$$

We estimate using Definition 3.5 for $\mathcal{L}^m$-a.e. $x \in \widetilde{W}^\pm$ that

$$\begin{aligned}&\phi(\bar{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x)) \\ &\qquad \le C_\phi \left(|\bar{u}_{\rho,r}(x) - f(x)| - |u(x) - f(x)|\right) |\bar{u}_{\rho,r}(x) - f(x)|^{p-1} \\ &\qquad \le C_\phi (1-\theta) |\gamma_{\rho,r\#} u(x) - u(x)| |\bar{u}_{\rho,r}(x) - f(x)|^{p-1} \\ &\qquad \le C |\gamma_{\rho,r\#} u(x) - u(x)|.\end{aligned}$$

In the final inequality we use the boundedness of u, or alternatively $p = 1$. We now employ Lemma 5.3 to estimate

$$\begin{aligned}\int_{\widetilde{W}_r^\pm} \phi(\bar{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x))\,dx &\le C \int_{\widetilde{W}_r^\pm} |\gamma_{\rho,r\#} u(x) - u(x)|\,dx \\ &\le C M_{\gamma_{\rho,r}^{-1}} |Du|(U_r^\pm).\end{aligned}$$

We have $M_{\gamma_{\rho,r}^{-1}} \le \rho r$. By the construction of Z_u, because $x_0 \in J_u \setminus Z_u$, choosing $r > 0$ small enough, we can enforce $|Du|(U_r^\pm) \le \epsilon r^{m-1}/(8C)$. Thus

$$\int_{\widetilde{W}^\pm} \phi(\bar{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x))\,dx \le \epsilon \rho r^m / 8. \tag{6.5}$$

It remains to estimate $\phi(\bar{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x))$ on $W_r^\pm$. By Definition 3.5, we have

$$\begin{aligned}-\phi(u(x) - f(x)) &\le -\phi(u_0(x) - f_0(x)) \\ &\qquad + C_\phi \big(|u_0(x) - f_0(x)| - |u(x) - f(x)|\big) |u_0(x) - f_0(x)|^{p-1}.\end{aligned}$$

Since $x \mapsto |u_0(x) - f_0(x)|^{p-1}$ is bounded, it follows for some constant $C > 0$ that

$$-\phi(u(x) - f(x)) \le -\phi(u_0(x) - f_0(x)) + C|u_0(x) - f_0(x) - u(x) + f(x)|. \tag{6.6}$$

We may estimate the final term in $W_r^\pm$ by

$$\begin{aligned}\int_{W_r^\pm} &|u_0(x) - f_0(x) - u(x) + f(x)|\,dx \\ &\le \int_{W_r^\pm} |u^\pm(x_0) - u(x)|\,dx + \int_{W_r^\pm} |\widetilde{f}(x_0) - f(x)|\,dx.\end{aligned} \tag{6.7}$$

We now recall the notation $u_x^z(t) = u(x+tz)$ for one-dimensional restrictions of functions of bounded variation in a direction z with $\|z\| = 1$, as well as the formula [3]

$$|\langle Du, z\rangle(A)| = \int_{P_z^\perp A} |Du_x^z|(\{t \in \mathbb{R} \mid x + tz \in A\})\,dx.$$

Writing $h_r = h_r^{(+)} - h_r^{(-)}$ for $h_r^{(+)}, h_r^{(-)} \ge 0$, we may then further estimate

$$\begin{aligned}\int_{W_r^\pm} &|u^\pm(x_0) - u(x)|\,dx \\ &= \int_{P_{z_\Gamma}^\perp U_r} \int_{f_\Gamma(v)}^{f_\Gamma(v) \pm \rho h_r^{(\pm)}(v)} |u^\pm(x_0) - u(v + tz_\Gamma)|\,dt\,dv \\ &= \int_{P_{z_\Gamma}^\perp U_r} \int_0^{\rho h_r^{(\pm)}(v)} |u^\pm(x_0) - u(g_\Gamma(v) \pm tz_\Gamma)|\,dt\,dv \\ &\le \int_{P_{z_\Gamma}^\perp U_r} \int_0^{\rho h_r^{(\pm)}(v)} |u^\pm(g_\Gamma(v)) - u(g_\Gamma(v) \pm tz_\Gamma)| + |u^\pm(g_\Gamma(v)) - u^\pm(x_0)|\,dt\,dv \\ &\le \int_{P_{z_\Gamma}^\perp U_r} \int_0^{\rho h_r^{(\pm)}(v)} |Du_{g_\Gamma(v)}^{z_\Gamma}|(\pm[0,t])\,dt\,dv + \rho r \int_{P_{z_\Gamma}^\perp U_r} |u^\pm(g_\Gamma(v)) - u^\pm(x_0)|\,dv \\ &\le \rho r \int_{P_{z_\Gamma}^\perp U_r} |Du_{g_\Gamma(v)}^{z_\Gamma}|(\pm[0,\rho r])\,dv + \rho r \int_{\Gamma \cap U_r} |u^\pm(x) - u^\pm(x_0)|\,dx \\ &\le \rho r |Du|(U_r^\pm) + \rho r \int_{\Gamma \cap U_r} |u^\pm(x) - u^\pm(x_0)|\,dx.\end{aligned}$$

In the semifinal step we have used on the second term the area formula on the transformation $g_\Gamma^{-1} = P_{z_\Gamma}^\perp$, observing that $\mathcal{J}_m P_{z_\Gamma}^\perp \le 1$. Since $x_0 \in J_u \setminus Z_u$ is a Lebesgue point of $u^\pm$ on Γ and $\Theta_{m-1}^*(|Du| \llcorner \Gamma^\pm; x_0) = 0$, choosing $r > 0$ small enough, we can make the final quantity less than $\rho r \cdot \epsilon r^{m-1}/(16C)$. This proves an estimate on the first term on the right hand side of (6.7). The second term we approximate by analogous arguments on $\widetilde{f}(x_0) - f$, using the fact that $x_0 \notin \widetilde{J}_f$ for the term $|Df|(U_r^\pm)$. Referring back to (6.6) and (6.7), we then deduce

$$-\int_{W_r^\pm} \phi(u(x) - f(x))\,dx \le -\int_{W_r^\pm} \phi(u_0(x) - f_0(x))\,dx + \epsilon \rho r^m/8. \tag{6.8}$$

It remains to estimate the transformed terms on $W_r^\pm$. Similarly to (6.6), using the essential boundedness of *both* u and f on U_r, or $p = 1$, we deduce for $\mathcal{L}^m$-a.e. $x \in W_r^\pm$

$$\phi(\bar{u}_{\rho,r}(x) - f(x)) \le \phi(\bar{u}_{\rho,r,0}(x) - f_0(x)) + C|\bar{u}_{\rho,r,0}(x) - f_0(x) - \bar{u}_{\rho,r}(x) + f(x)|. \tag{6.9}$$

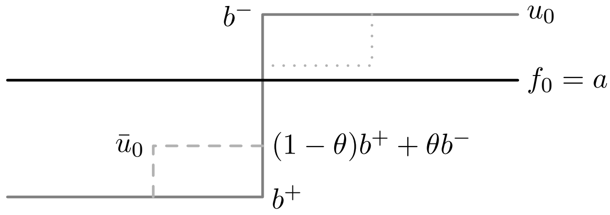

FIGURE 6.2. *The situation in the proof of Lemma 6.3. Here $\bar u_0 = \bar u_{\rho,r,0}$ for $\rho > 0$ (dashed line). The dotted ine indicates the alternative $\bar u_{-\rho,r,0}$ that we may need to use if the curvature of J_u is high at x_0, as measured by the double-Lipschitz comparison condition for R.*

A simple application of the area formula and repeating the arguments for the untransformed term above, yields again the estimate

$$\int_{W_r^\pm} \phi(\bar u_{\rho,r}(x) - f(x))\,dx \le \int_{W_r^\pm} \phi(\bar u_{\rho,r,0}(x) - f_0(x))\,dx + \epsilon\rho r^m/8. \tag{6.10}$$

Summing the estimates (6.4), (6.5), (6.8) and (6.10), and minding that $\bar u_{\rho,r} = u$ outside U_r, we deduce (6.3). □

LEMMA 6.3. *Suppose ϕ is p-increasing with $p > 1$. Then there exist $\theta \in (0,1)$ and a constant $C = C(\phi, u^\pm(x_0), \widetilde f(x_0)) > 0$ such that for $\rho > 0$ holds*

$$\int_{U_r} \phi(\bar u_{\rho,r,0}(x) - f_0(x))\,dx + \int_{U_r} \phi(\bar u_{-\rho,r,0}(x) - f_0(x))\,dx \\ - 2\int_{U_r} \phi(u_0(x) - f_0(x))\,dx \le -C\rho r^m.$$

Proof. Let $\rho > 0$. Observe that outside $\widehat W_r^\pm := U_r^\pm \cap \gamma_{\pm\rho,r}(U_r^\mp)$ the transformed piecewise constant functions agree with the originals,

$$\bar u_{\pm\rho,r,0}(x) = u_0(x), \quad (x \in U_r \setminus \widehat W_r^\pm).$$

Letting

$$K^\pm := \operatorname*{ess\,sup}_{x\in \widehat W_r^\pm} \bigl(|\bar u_{\pm\rho,r,0}(x) - f_0(x)| - |u_0(x) - f_0(x)|\bigr)$$

and using Definition 3.5, we derive for $x \in \widehat W_r^\pm$ the estimate

$$\begin{aligned}\phi(\bar u_{\pm\rho,r,0}(x) - f_0(x)) - \phi(u_0(x) - f_0(x)) \\ &\le C_\phi\bigl(|\bar u_{\pm\rho,r,0}(x) - f_0(x)| - |u_0(x) - f_0(x)|\bigr)|\bar u_{\pm\rho,r,0}(x) - f_0(x)|^{p-1} \\ &\le K^\pm C_\phi |\bar u_{\pm\rho,r,0}(x) - f_0(x)|^{p-1}.\end{aligned}$$

Overall we then have

$$\int_{U_r} \phi(\bar u_{\pm\rho,r,0}(x) - f_0(x))\,dx - \int_{U_r} \phi(u_0(x) - f_0(x))\,dx \\ \le K^\pm C_\phi \int_{\widehat W_r^\pm} |\bar u_{\pm\rho,r,0}(x) - f_0(x)|^{p-1}\,dx. \tag{6.11}$$

Let us denote $b^\pm := u^\pm(x_0)$ and $a := \widetilde{f}(x_0)$. Then

$$|\bar{u}_{\rho,r,0}(x) - f_0(x)|^{p-1} = \zeta^+ := |(1-\theta)(b^- - a) + \theta(b^+ - a)|^{p-1}, \quad (x \in \widehat{W}_r^+), \quad \text{and}$$
$$|\bar{u}_{-\rho,r,0}(x) - f_0(x)|^{p-1} = \zeta^- := |(1-\theta)(b^+ - a) + \theta(b^- - a)|^{p-1}, \quad (x \in \widehat{W}_r^-).$$

Moreover,

$$K^+ = |(1-\theta)(b^- - a) + \theta(b^+ - a)| - |b^+ - a|, \quad \text{and}$$
$$K^- = |(1-\theta)(b^+ - a) + \theta(b^- - a)| - |b^- - a|.$$

By Lemma A.1 in the appendix

$$\mathcal{L}^m(\widehat{W}_r^+) = \mathcal{L}^m(\widehat{W}_r^-) = \int_{V_\Gamma} \rho |h_r(v)| \, d\mathcal{H}^{m-1}(v) = C_h \rho r^m$$

for some constant $C = C_h$. Thus summing (6.11) for $\pm\rho$ gives

$$\begin{aligned}\int_{U_r} \phi(\bar{u}_{\rho,r,0}(x) - f_0(x)) \, dx + \int_{U_r} \phi(\bar{u}_{-\rho,r,0}(x) - f_0(x)) \, dx \\ - 2\int_{U_r} \phi(u_0(x) - f_0(x)) \, dx \\ \leq C_\phi C_h \rho r^m (K^+\zeta^+ + K^-\zeta^-).\end{aligned}$$

In order to reach our conclusion, it therefore remains to show that

$$K^+\zeta^+ + K^-\zeta^- < 0. \tag{6.12}$$

We calculate

$$K^+ \leq (1-\theta)(|b^- - a| - |b^+ - a|), \tag{6.13}$$

and

$$K^- \leq (1-\theta)(|b^+ - a| - |b^- - a|). \tag{6.14}$$

Therefore

$$K^+\zeta^+ + K^-\zeta^- \leq (1-\theta)(|b^- - a| - |b^+ - a|)(\zeta^+ - \zeta^-).$$

We concentrate on the case $b^+ < b^-$ with $a \geq (b^+ + b^-)/2$, other cases handled analogously by appropriate changes of roles and negations. This is illustrated in Figure 6.2. Then

$$|b^+ - a| \geq |b^- - a|. \tag{6.15}$$

If (6.15) holds strictly, we have $\zeta^+ > \zeta^-$ for $\theta \in (0,1)$ large enough. This shows (6.12) and is the only place where we need the assumption $p > 1$. If (6.15) does not hold strictly, i.e., $a = (b^+ + b^-)/2$, we may have $\zeta^+ = \zeta^-$, but observe that $K^+ = 0$ and (6.13) holds strictly for large θ. Indeed, this is the case whenever $b^+ < a, b^-$, because some interpolation $(1-\theta)b^- + \theta b^+$ is always closer to a than b^+ is. We have thus proved (6.12), and may conclude the proof of the lemma. □

Lemma 6.4. *Suppose ϕ is p-increasing with $1 < p < \infty$, and* both $u, f \in \mathrm{BV}(\Omega) \cap L^\infty_{\mathrm{loc}}(\Omega)$. *Let $x_0 \in J_u \setminus (\widetilde{J}_f \cup S_f \cup Z_u)$. Then there exist $\theta \in (0,1)$, $r_0 > 0$, independent of ρ, and a constant $C = C(\phi, u^\pm(x_0), \widetilde{f}(x_0)) > 0$, such that whenever $0 < r < r_0$ and $0 < \rho < 1$, it holds*

$$\int_\Omega \phi(\bar{u}_{\rho,r}(x) - f(x))\,dx + \int_\Omega \phi(\bar{u}_{-\rho,r}(x) - f(x))\,dx - 2\int_\Omega \phi(u(x) - f(x))\,dx \le -C\rho r^m. \tag{6.16}$$

Proof. Combine Lemma 6.2 and Lemma 6.3, choosing $\epsilon > 0$ small enough in the former. □

6.3. The effect of the shift transformation on the regulariser. We now summarise the estimates we get for the regulariser R using double-Lipschitz comparability and the Lipschitz transformations of §5.

Lemma 6.5. *Suppose R is a double-Lipschitz comparable and $x_0 \in J_u \setminus Z_u$. Then there exists a constant $C = C(u, x_0)$ and $r_0 > 0$ such that for $0 < r < r_0$ and $0 < \rho < 1$ holds*

$$R(\bar{u}_{\rho,r}) + R(\bar{u}_{-\rho,r}) - 2R(u) \le C\rho^2 r^{m-1}.$$

Proof. We know from Lemma 5.2 the existence of a constant $C > 0$ such that $T_{\gamma_{\rho,r},\gamma_{-\rho,r}}$ defined in (3.2e) satisfies

$$T_{\gamma_{\rho,r},\gamma_{-\rho,r}} \le C\rho^2.$$

By convexity

$$R(\bar{u}_{\rho,r}) + R(\bar{u}_{-\rho,r}) - 2R(u) \le (1-\theta)\big(R(\gamma_{\rho,r\#}u) + R(\gamma_{-\rho,r\#}u) - 2R(u)\big).$$

By the double-Lipschitz comparability of R, we thus find that

$$R(\bar{u}_{\rho,r}) + R(\bar{u}_{-\rho,r}) - 2R(u) \le (1-\theta)R^a T_{\gamma_{\rho,r},\gamma_{-\rho,r}} |Du|(\mathrm{cl}\, U_r).$$

Since $x_0 \in J_u \setminus Z_u$, whenever $r > 0$ is small enough, we have for a constant $C' > 0$ that

$$|Du|(\mathrm{cl}\, U_r)) \le |Du|(B(x_0, C'r)) \le 2\omega_{m-1}|u^+(x_0) - u^-(x_0)|(C'r)^{m-1}.$$

The claim follows by combining the above estimates. □

6.4. Patching it all together. We may finally prove our main result for $p > 1$ by combining the above lemmas as follows.

Proof of Theorem 3.7. Suppose, to reach a contradiction, that $\mathcal{H}^{m-1}(J_u \setminus J_f) > 0$. Since $\mathcal{H}^{m-1}(Z_u) = 0$, $\mathcal{H}^{m-1}(S_f \setminus J_f) = 0$ with $J_f \subset S_f$, and by Proposition 2.1, $\mathcal{H}^{m-1}(\widetilde{J}_f \setminus J_f) = 0$ with $J_f \subset \widetilde{J}_f$, we may select a point $x_0 \in J_u \setminus (\widetilde{J}_f \cup S_f \cup Z_u)$. We then apply Lemma 6.4 for the fidelity estimate

$$\begin{aligned}\int_\Omega \phi(\bar{u}_{\rho,r}(x) - f(x))\,dx + \int_\Omega \phi(\bar{u}_{-\rho,r}(x) - f(x))\,dx \\ - 2\int_\Omega \phi(u(x) - f(x))\,dx \le -C_1\rho r^m,\end{aligned} \tag{6.17}$$

where $0 < r < r_1$ and $0 < \rho < 1$. Lemma 6.5 gives for the regulariser and $0 < r < r_2$ the estimate

$$R(\bar{u}_{\rho,r}) - R(\bar{u}_{-\rho,r}) - 2R(u) \le C_2\rho^2 r^{m-1}.$$

All the constants $C_1, C_2, r_1, r_2 > 0$ are independent of $\rho \in (0,1)$. Picking $0 < r < \min\{r_1, r_2\}$, and summing these estimates, we obtain

$$\begin{aligned}\left(\int_\Omega \phi(\bar{u}_{\rho,r}(x) - f(x))\,dx + R(\bar{u}_{\rho,r})\right) + \left(\int_\Omega \phi(\bar{u}_{-\rho,r}(x) - f(x))\,dx + R(\bar{u}_{-\rho,r})\right)&\\ - 2\left(\int_\Omega \phi(u(x) - f(x))\,dx + R(u)\right)&\\ \le C_2\rho^2 r^{m-1} - C_1\rho r^m.&\end{aligned}$$

If $\rho > 0$ is small enough, this is negative. Thus either

$$\int_\Omega \phi(\bar{u}_{\rho,r}(x) - f(x))\,dx + R(\bar{u}_{\rho,r}) < \int_\Omega \phi(u(x) - f(x))\,dx + R(u), \quad \text{or}$$
$$\int_\Omega \phi(\bar{u}_{-\rho,r}(x) - f(x))\,dx + R(\bar{u}_{-\rho,r}) < \int_\Omega \phi(u(x) - f(x))\,dx + R(u).$$

This contradicts the optimality of u. Therefore $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$. □

7. Considerations for the L^1 fidelity. The only difficulty in extending the proofs of §6 for p-increasing ($p > 1$) fidelities ϕ to the L^1 fidelity $\phi(x) = x$, is Lemma 6.3. We do not necessarily get a constant $C > 0$ there, but $C = 0$. Then the argument in the proof of Theorem 3.7 does not go through. This has the consequence that it is possible to have $\mathcal{H}^{m-1}(J_u \setminus J_f) > 0$. However, as we will see in this section, the residual $J_u \setminus J_f$ has a regular structure, although our result is somewhat weaker than the known result for TV [18].

In this section, we still define h_r as before, as

$$h_r(v) := r\bar{h}((v - \mathfrak{v}_{x_0})/r),$$

for some $0 \neq \bar{h} \in W_0^{1,\infty}(z_\Gamma^\perp \cap B(0,1))$, with $0 \le \bar{h} \le 1$ unless explicitly stated otherwise. However, we also denote

$$h_r^{v_0}(v) := r\bar{h}((v - v_0)/r),$$

when we want to be more explicit about the base point. Noting that $I_r > 0$ under these assumptions, we denote for brevity

$$I_r := \int_{V_\Gamma} h_r\,dv,$$

and

$$\lambda_u(x) := |u^+(x) - u^-(x)|.$$

Finally, we need the missing technical curvature definition for Theorem 3.8.

Definition 7.1. *Let R be an admissible regularisation functional on* BV(Ω). *We define the* transformation differential *of R at u as*

$$\mathcal{D}_u^R(\gamma) := \lim_{\rho \searrow 0} \frac{R((\rho\gamma + (1-\rho)\iota)_\# u) - R(u)}{\rho}, \quad (\gamma \in \mathcal{F}(\Omega)),$$

when the limit exists. With $\bar{h} \in W_0^{1,\infty}(z_\Gamma^\perp \cap B(0,1))$*, we then define the* pointwise R-curvature *at* x_0 *along the Lipschitz graph* Γ*, if the limit exists, as*

$$\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) := \lim_{r \searrow 0} I_r^{-1} \mathcal{D}_R(\gamma_{h_r}), \quad (x_0 \in \Gamma).$$

Since our results do not depend on the choice of $\bar{h}$*, we simply write* $\mathcal{C}_u^{R,\Gamma}(x_0) := \mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0)$ *in the statement of Theorem 3.8.*

Remark 7.1. In many ways, it would make more sense to define the pointwise curvature by

$$\widetilde{\mathcal{C}}_u^{R,\Gamma,\bar{h}}(x_0) := \lim_{r \searrow 0} \frac{\omega_{m-1} \mathcal{D}_u^R(\gamma_{1,r})}{I_1 r |Du|(B(x_0,r))}.$$

At a point $x_0 \in J_u$ this would give the normalised curvature $\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0)/\lambda_u(x_0)$; see Lemma B.2. For practical reasons, in order to be able to work easily at points $x_0 \in \Gamma \setminus J_u$, we use the earlier definition, however.

7.1. Expressions for the pointwise R-curvature. The proofs of the curvature condition in Theorem 3.8 is mainly based on the expressions in the following two lemmas, plus the regularity results that follows. Since the proofs of these two lemmas are mainly technical and not very informative, we relegate them to Appendix B.

The first one of the lemmas is our rough counterpart of Lemma 5.4 for $p > 1$. The proof uses a different technique, based on only one-sided Lipschitz estimates.

Lemma 7.2. *Suppose u solves (P). Let $\Gamma \subset \Omega$ be a Lipschitz $(m-1)$-graph. Then*

$$\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) = \lambda_u(x_0) C_\phi, \quad (\mathcal{H}^{m-1}\text{-a.e. } x_0 \in (J_u \setminus J_f) \cap \Gamma).$$

For the second one of the curvature lemmas, giving a more familar mean curvature expression for the pointwise R-curvature, we require our rather strong assumption that Du is "essentially piecewise constant" around $J_f \setminus J_u$.

Lemma 7.3. *Let $\Gamma \subset \Omega$ be a Lipschitz $(m-1)$-graph. At $\mathcal{H}^{m-1}$-a.e. point $x_0 \in \Gamma$, if $\Theta_m(|Du| \llcorner \Omega \setminus \Gamma; x_0) = 0$, then*

$$\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) = -\operatorname{div} \frac{\nabla f_\Gamma(P_{z_\Gamma}^\perp x_0)}{\sqrt{1 + \|\nabla f_\Gamma(P_{z_\Gamma}^\perp x_0)\|^2}} \cdot R^s \lambda_u(x_0).$$

7.2. Regularity. The following result will allow us to obtain additional regularity from the expression for the pointwise R-curvature in Lemma 7.3.

Lemma 7.4. *Let $f \in W^{1,\infty}(V)$, and $0 \le \lambda \in L^\infty(V)$, where $V \subset z^\perp \cap B(0,r_0)$, $z \in \mathbb{R}^m$, $r_0 > 0$. Suppose for some $\alpha > 0$ we have*

$$-\operatorname{div} \frac{\nabla f(v)}{\sqrt{1 + \|\nabla f(v)\|^2}} = \alpha^{-1}, \tag{7.1}$$

for $\mathcal{H}^{m-1}$-a.e. $v \in V$. Then (7.1) holds for every $v \in V$, and $f \in C^{2,\cap}(V)$.

Proof. Since (7.1) holds almost everywhere, also

$$\int_V \left\langle \frac{\nabla f(v)}{\sqrt{1 + \|\nabla f(v)\|^2}}, \nabla h(v) \right\rangle dv = \alpha^{-1} \int_V h(v)\, dv, \quad (h \in W_0^{1,\infty}(V)). \tag{7.2}$$

After showing C^2 regularity of f, (7.1) then holds for all $v \in V$. To show this, let us start by defining

$$\mathcal{J}(q) := \int_V \sqrt{1 + \|\nabla q(x)\|^2}\, dx - \alpha^{-1} \int_V q(x) - f(x)\, dx,$$

and consider the problem

$$\min\{\mathcal{J}(q) \mid q \in Q\}, \tag{7.3}$$

where the domain

$$Q := \{q \in W^{1,\infty}(V) \mid q|\partial V = f|\partial V\}.$$

Since $t \mapsto \sqrt{1+t^2}$ is strictly convex, and the weak solution $h \in W_0^{1,\infty}(V)$ to the differential equation

$$\nabla h = \nabla \widetilde{h} \text{ on } V, \quad h|\partial V = 0,$$

is unique for any $\widetilde{h} \in W_0^{1,\infty}(V)$, we find that $\mathcal{J}$ is strictly convex on Q. It therefore has a unique minimiser. But (7.2) is exactly the necessary and sufficient optimality condition for f to be a minimiser of $\mathcal{J}$. This can be deduced from the following paragraphs studying a slightly modified functional $\widehat{\mathcal{J}}$. At this moment it is important to notice that f must then be the unique minimiser of $\mathcal{J}$.

We intend to use the regularity results on quasilinear elliptic partial differential equations to show that f satisfies the claimed regularity properties. To do this, we however need to force some coercivity/ellipticity properties. We therefore define

$$\widehat{\mathcal{J}}(q) := \int_V g(\nabla q(x))\, dx - \alpha^{-1} \int_V \big(u(x) - f(x)\big)\, dx,$$

where

$$g(p) := \sqrt{1 + \|p\|^2} + \sum_{i=1}^m \psi(|p_i|),$$

and

$$\psi(t) = C_0\big(\max\{t - K, 0\}\big)^4$$

for $K := 2\sup_{x \in V} \|\nabla f(x)\|$, and yet undetermined $C_0 > 0$. Observe that $\psi \in C^3(\mathbb{R})$. Now $\mathcal{J} \le \widehat{\mathcal{J}}$ and $\widehat{\mathcal{J}}(f) = \mathcal{J}(f)$, so that $\widehat{\mathcal{J}}$ also has the unique minimiser f within U.

We now write the optimality conditions for potential minimisers $q \in Q$ of $\widehat{\mathcal{J}}$. As differentials of the mappings

$$\rho \mapsto \widehat{\mathcal{J}}(q + \rho h), \quad (h \in W_0^{1,\infty}(V)),$$

we calculate

$$\int_V \langle \nabla g(\nabla(q + \rho h)(x)), \nabla h(x)\rangle\, dx - \alpha^{-1} \int_V h(x)\, dx.$$

If q minimises $\widehat{\mathcal{J}}$, we obtain by setting $\rho = 0$ the optimality condition

$$\int_V \langle \nabla g(\nabla q(x)), \nabla h(x)\rangle\, dx = \alpha^{-1} \int_V h(x)\, dx, \quad (h \in W_0^{1,\infty}(V)),$$

where

$$\nabla g(p) = \frac{p}{\sqrt{1+\|p\|^2}} + \sum_{i=1}^{m} \psi'(|p_i|)(\operatorname{sgn} p_i)e_i.$$

It follows that q is a solution of the quasilinear elliptic differential equation

$$\operatorname{div} A(\nabla q) = -\alpha^{-1}, \quad q \in W^{1,\infty}(V), \quad q|\partial V = f|\partial V, \tag{7.4}$$

where

$$A(p) = (A_1(p), \ldots, A_m(p)) := \nabla g(p).$$

We want to show that (7.4) has a solution $\widehat{q} \in C^{2,\lambda}(V)$ for any $\lambda \in (0,1)$. But $\widehat{\mathcal{J}}$ had the unique minimiser f. Therefore $f = \widehat{q} \in C^{2,\lambda}(V)$ for any $\lambda \in (0,1)$. From (7.4) and the definition of ψ we have moreover that (7.1) holds.

To show the existence of a solution $\widehat{q} \in C^{2,\lambda}(V)$, we employ [20, Theorem 15.19]. To do so, we need to show that

$$A_i \in C^{1,\lambda}(\mathbb{R}^m), \quad (i = 1, \ldots, m), \tag{7.5}$$

that for some $\tau > -1$ and $C_1, C_2 \in \mathbb{R}$ both

$$\langle p, A(p)\rangle \geq \|p\|^{2+\tau} - C_1, \tag{7.6}$$

and

$$\|\nabla A(p)\| \leq C_2(1 + \|p\|)^{\tau}. \tag{7.7}$$

It is to force these conditions, why we introduced the ψ-penalty in g. Minding that $\psi \in C^3(\mathbb{R})$, we have $A \in C^2(\mathbb{R}^m; \mathbb{R}^m)$, whence condition (7.5) readily follows. Moreover, we have for some $C_0, C_1 \geq 0$ that

$$\sum_{i=1}^{m} \psi'(p_i)|p_i| \geq C_0 \sum_{i=1}^{m} |p_i|^4 - C_1 \geq \|p\|^4 - C_1,$$

so that

$$\langle p, A(p)\rangle = \frac{\|p\|^2}{\sqrt{1+\|p\|^2}} + \sum_{i=1}^{m} \psi'(p_i)|p_i| \geq \|p\|^4 - C_1.$$

Thus (7.6) holds with $\tau = 2$. It remains to show that (7.7) also holds with $\tau = 2$. We have

$$\nabla A(p) = \frac{I}{\sqrt{1+\|p\|^2}} - \frac{p \otimes p}{(1+\|p\|^2)^{3/2}} + \sum_{i=1}^{m} \psi''(|p_i|)e_i \otimes e_i.$$

The first two terms are bounded, while for for some $C_3, C_4 > 0$ we have the estimate

$$|\psi''(|p_i|)| \leq C_3 + C_4 p_i^2.$$

Therefore, for some $C_5, C_2 > 0$, we find that

$$\|\nabla A(p)\| \leq C_5 + C_4\|p\|^2 \leq C_2(1 + \|p\|)^2.$$

This proves (7.7), concluding the proof. □

7.3. Proof of the main result for $p = 1$. We may now prove our main result for 1-increasing fidelities by combining the results of the above lemmas.

Proof of Theorem 3.8. Let $\{\Gamma_i\}_{i=1}^\infty$ be the collection of Lipschitz graphs from (6.1). Repeated application of Lemma 7.2 for each $\Gamma = \Gamma_i$, $(i \in \mathbb{Z}^+)$ shows that

$$\mathcal{C}_u^{R,\Gamma_i,\bar{h}}(x_0) = \lambda_u(x_0)C_\phi \quad \text{for} \quad \mathcal{H}^{m-1}\text{-a.e. } x_0 \in (J_u \setminus J_f) \cap \Gamma_i. \tag{7.8}$$

Choosing $\Lambda_i = \Gamma_i$ proves (3.4).

It remains to prove the higher regularity (3.5) under assumptions of approximate piecewise constancy. We fix $\Gamma = \Gamma_i$ for some $i \in \mathbb{Z}^+$. By assumption $\Theta_m(|Du|\llcorner\Omega \setminus \Gamma; x_0) = 0$ at $\mathcal{H}^{m-1}$-a.e. $x_0 \in \Gamma$. By Lemma 7.3 we therefore have at $\mathcal{H}^{m-1}$-a.e. $x_0 \in \Gamma$ the expression

$$\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) = -\operatorname{div} \frac{\nabla f_\Gamma(P_{z_\Gamma}^\perp x_0)}{\sqrt{1 + \|\nabla f_\Gamma(P_{z_\Gamma}^\perp x_0)\|^2}} \cdot \lambda_u(x_0).$$

We plan to use Lemma 7.4. This depends on the domain V_Γ being open, and (7.8) holding $\mathcal{H}^{m-1}$-a.e. on $\Gamma = \Gamma_i$, instead of just on $(J_u \setminus J_f) \cap \Gamma_i$. We therefore need a tiny extra argument before the application of the lemma.

We recollect from the proof of Lemma 7.3 the expression

$$\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) = \lim_{r \searrow 0} c_{1,f}(h_r/I_r) \cdot \lambda_u(x_0)$$

for

$$c_{1,f}(h) := \int_V \left\langle \frac{\nabla f_\Gamma(v)}{\sqrt{1 + \|\nabla f_\Gamma(v)\|^2}}, \nabla h(v) \right\rangle dv.$$

The functional $c_{1,f}$ is continuous on $H_0^1(V_\Gamma)$. From this, it is not difficult to see that each of the sets

$$W_{k,r} := \{v \in V_\Gamma \mid |c_{1,f}(h_r^v/I_r) - C_\phi| \geq 1/k\}, \quad (k \in \mathbb{Z}^+,\ r > 0).$$

is closed within V_Γ. Therefore so are the sets

$$W_k^\sigma := \bigcap_{0<r<\sigma} W_{k,r}, \quad (k \in \mathbb{Z}^+,\ \sigma > 0),$$

consisting of points $v \in V_\Gamma$ where

$$\liminf_{r \searrow 0} |c_{1,f}(h_r^v/I_r) - C_\phi| \geq 1/k.$$

Suppose $\operatorname{ri} W_k^\sigma \neq \emptyset$. However $\mathcal{H}^{m-1}(\operatorname{ri} W_k^\sigma \cap (J_u \setminus J_f)) = 0$, because otherwise we would find a set of positive mass of points $x_0 \in J_u \setminus J_f$ where $\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) \neq \lambda_u(x_0)C_\phi$, and reach a contradiction to (7.8). Because we could always replace Γ by $\Gamma \setminus g_\Gamma(\operatorname{cl}\operatorname{ri} W_k^\sigma)$, we may therefore assume that $\operatorname{ri} W_k^\sigma = \emptyset$ for every $k \in \mathbb{Z}^+$ and $\sigma > 0$. In particular every W_k^σ is nowhere dense in V_Γ. The Baire category theorem then says that

$$V'_\Gamma := \bigcap_{k=1}^\infty \bigcap_{j=1}^\infty (V_\Gamma \setminus W_k^{1/j}) = \bigcap_{k=1}^\infty \bigcap_{j=1}^\infty \bigcup_{0<r<1/j} (V_\Gamma \setminus W_{k,r}).$$

is a dense open set within V_Γ. But every $v \in V'_\Gamma$ satisfies for some $r_j \searrow 0$ the limit

$$\lim_{j\to\infty} c_{1,f}(h^v_{r_j}/I_{r_j}) = C_\phi.$$

Considering that $\lim_{r\searrow 0} c_{1,f}(h^v_r/I_r)$ exists at a Lebesgue point of ∇f_Γ, we deduce that $\mathcal{C}^{R,\Gamma,\bar h}_u(g_\Gamma(v)) = \lambda_u(g_\Gamma(v))C_\phi$ for $\mathcal{H}^{m-1}$-a.e. $v \in V'_\Gamma$. Moreover, as required

$$\mathcal{H}^{m-1}(J_u \setminus (J_f \cup \Lambda)) = 0$$

for $\Lambda = \bigcup_{i=1}^\infty \Lambda_i$ with $\Lambda_i := g_\Gamma(V'_{\Gamma_i})$. Finally, we refer to Lemma 7.4 (on a small ball around any given point $v \in V_{\Lambda_i}$) to conclude the regularity of Λ_i under the assumption that $\Theta_m(|Du|\llcorner\Omega \setminus \Lambda_i; x) = 0$ at $\mathcal{H}^{m-1}$-a.e. point $x \in \Lambda_i$. □

Acknowledgements. This manuscript has been prepared over the course of several years, on the side of "paying work". While the author was at the Institute for Mathematics and Scientific Computing at the University of Graz, this work was financially supported by the SFB research program F32 "Mathematical Optimization and Applications in Biomedical Sciences" of the Austrian Science Fund (FWF). While at the Department of Applied Mathematics and Theoretical Physics at the University of Cambridge, this work was financially supported by the King Abdullah University of Science and Technology (KAUST) Award No. KUK-I1-007-43 as well as the EPSRC / Isaac Newton Trust Small Grant "Non-smooth geometric reconstruction for high resolution MRI imaging of fluid transport in bed reactors", and the EPSRC first grant Nr. EP/J009539/1 "Sparse & Higher-order Image Restoration". At the Research Center on Mathematical Modeling (Modemat) at the Escuela Politécnica Nacional de Quito, this work has been supported by the Prometeo initiative of the Senescyt.

The author would like to express sincere gratitude to Simon Morgan, Antonin Chambolle, Kristian Bredies, and Carola-Bibiane Schönlieb for fruitful discussions. Moreover, the author is grateful to Kostas Papafitsoros for an additional proofreading of the manuscript.

Appendix A. Areas and lengths under Lipschitz transformations.

We state and prove the following well-known expressions for completeness.

Lemma A.1. *Let $\Gamma \subset \mathbb{R}^m$ be a Lipschitz $(m-1)$-graph. Suppose $x_0 \in \Gamma$ and $\delta_0 > 0$ are such that $B(g_\Gamma^{-1}(x_0), \delta_0) \subset V_\Gamma$. Let $\lambda \in L^1(\Gamma)$ and $0 \le h \in W^{1,\infty}_0(V_\Gamma)$, with $\|h\|_{L^1(V_\Gamma)} > 0$. Define*

$$\Gamma_\rho := \gamma_{\rho h}(\Gamma) = \{g_\Gamma(v) + \rho h(v) z_\Gamma \mid v \in V_\Gamma\}, \quad (\rho \in \mathbb{R}).$$

Then

$$\mathcal{L}^m(\Gamma^-_\rho \cap \Gamma^+) + \mathcal{L}^m(\Gamma^+_\rho \cap \Gamma^-) = |\rho| \int_{V_\Gamma} h(v)\,dv, \tag{A.1}$$

$$\int_\Gamma \lambda(x)\,d\mathcal{H}^{m-1}(x) = \int_{V_\Gamma} \lambda(g_\Gamma(v))\sqrt{1 + \|\nabla f_\Gamma(v)\|^2}\,dv, \quad \textit{and} \tag{A.2}$$

$$\int_{\Gamma_\rho} \lambda(x)\,d\mathcal{H}^{m-1}(x) = \int_{V_\Gamma\cap} \lambda(g_\Gamma(v))\sqrt{1 + \|\nabla f_\Gamma(v) + \rho\nabla h(v)\|^2}\,dv. \tag{A.3}$$

Proof. Without loss of generality, we may assume that $x_0 = 0$, and $g_\Gamma^{-1}(x_0) = 0$. We thus calculate for $\rho \ge 0$ that

$$\Gamma^-_\rho \cap \Gamma^+ = \{g_\Gamma(v) + \tau z_\Gamma \mid v \in V_\Gamma,\ 0 < \tau < |\rho| h(v)\}.$$

Analogously for $\rho < 0$ we obtain

$$\Gamma_\rho^- \cap \Gamma^+ = \{g_\Gamma(v) - \tau z_\Gamma \mid v \in V_\Gamma,\ 0 < \tau < |\rho| h(v)\}.$$

Application of Fubini's theorem therefore confirms (A.1).

To prove (A.2), we write

$$\Gamma = \{g_\Gamma(v) \mid v \in V_\Gamma\}.$$

Thus the area formula (2.2) gives

$$\int_\Gamma \lambda(x)\, d\mathcal{H}^{m-1}(x) = \int_{V_\Gamma} \lambda(g_\Gamma(v)) \mathcal{J}_{m-1} g_\Gamma(v)\, dv.$$

Writing

$$g_\Gamma(v) = v + f_\Gamma(v) z_\Gamma$$

we have

$$\nabla g_\Gamma(v) = H^* + \nabla f_\Gamma(v) \otimes z_\Gamma,$$

where $H \in \mathcal{L}(V_\Gamma; \mathbb{R}^m)$ is the embedding operator $H(v) = v$. We have $H^* \circ H = I$ and $H^* z_\Gamma = 0$. Consequently

$$\nabla g_\Gamma(v) \circ [\nabla g_\Gamma(v)]^* = I + \nabla f_\Gamma(v) \otimes \nabla f_\Gamma(v),$$

The eigenvalue of $\nabla g_\Gamma(v)^* \circ \nabla g_\Gamma(v)$. corresponding to eigenvector $\nabla f_\Gamma(v)$ is $1 + \|\nabla f_\Gamma(v)\|^2$, while the other eigenvalues equal 1. Thus

$$\mathcal{J}_{m-1} g_\Gamma(v) = \sqrt{\det(\nabla g_\Gamma(v) \circ [\nabla g_\Gamma(v)]^*)} = \sqrt{1 + \|\nabla f_\Gamma(v)\|^2},$$

giving the well-known formula for $\mathcal{H}^{m-1}(\Gamma \setminus \Gamma_\rho)$ and confirming (A.2).

The same arguments prove also (A.3). Indeed, we have

$$\Gamma_\rho = \{q_\rho(v) \mid v \in V_\Gamma\}.$$

where

$$q_\rho(v) := v + (f_\Gamma + h)(v) z_\Gamma,$$

so the area formula (2.2) again gives

$$\int_{\Gamma_\rho} \lambda(x)\, d\mathcal{H}^{m-1}(x) = \int_{V_\Gamma} \lambda(g_\Gamma(v)) \mathcal{J}_{m-1} q_\rho(v)\, dv,$$

where

$$\mathcal{J}_{m-1} q_\rho(v) = \sqrt{1 + \|\nabla f_\Gamma(v) + \rho \nabla h_\Gamma(v)\|^2}. \qquad \square$$

Appendix B. Proofs for $p = 1$ (§7).

In this appendix, we provide the proofs for some of the lemmas required for the $p = 1$ case in §7. We begin with a modified version of the idelity estimate Lemma 6.4 in Appendix B.1. Observe that this version does not require the L^∞_{loc} bounds. Employing this result, we then prove Lemma 7.2 and Lemma 7.3, mainly concerning the regulariser, in Appendix B.2 and Appendix B.3 respectively.

B.1. Estimate of the fidelity. LEMMA B.1. *Suppose $\phi(t)$ is 1-increasing, and that $u, f \in \mathrm{BV}(\Omega)$. Let $x_0 \in J_u \setminus (S_f \cup Z_u)$. Given $\epsilon > 0$, there exist $\sigma \in \{-1, +1\}$, $r_0 > 0$ and $\theta \in [0,1]$, such that for any $r \in (0, r_0)$, $-1 < \rho < 1$, and $\bar{h} \in W_0^{1,\infty}(z_\Gamma^\perp \cap B(0,1))$ with $-1 \le \bar{h} \le 1$, we have*

$$\begin{aligned}\int_\Omega \phi(\bar{u}_{\rho,r}(x) - f(x))\,dx - \int_\Omega \phi(u(x) - f(x))\,dx \\ \le \epsilon \rho r^m - \sigma(1-\theta)\rho C_\phi \lambda_u(x_0) I_r, \quad (0 < r < r_0, -1 < \rho < 1).\end{aligned} \tag{B.1}$$

Proof. We assume that $\rho \ge 0$; the case $\rho < 0$ can be handled by negating $\bar{h}$. Using Lemma 6.2 to pass from the functions $(\bar{u}_{\rho,r}, u, f)$ to $(\bar{u}_{\rho,r,0}, u_0, f_0)$, it suffices to prove the estimate corresponding to (B.1) for the latter piecewise constant functions.

Referring to Definition 3.5 we deduce

$$\begin{aligned}&\phi(\bar{u}_{\rho,r,0}(x) - f_0(x)) - \phi(u_0(x) - f_0(x)) \\ &\qquad \le C_\phi(|\bar{u}_{\rho,r,0}(x) - f_0(x)| - |u_0(x) - f_0(x)|)|\bar{u}_{\rho,r,0}(x) - f_0(x)|^0.\end{aligned}$$

This is non-zero only in $W_r^+ := U_r^+ \cap \gamma_{\rho,r}(U_r^-)$, and in $W_r^- := U_r^- \cap \gamma_{\rho,r}(U_r^+)$. For $x \in W_r^+$, as in the proof of Lemma 6.3,

$$\begin{aligned}K^+ &:= |(1-\theta)(b^- - a) + \theta(b^+ - a)| - |b^+ - a| \\ &= |\bar{u}_{\rho,r,0}(x) - f_0(x)| - |u_0(x) - f_0(x)|.\end{aligned}$$

Here we again denote $b^\pm := u^\pm(x_0)$ and $a := \widetilde{f}(x_0)$. Also

$$\zeta^+ := |(1-\theta)(b^- - a) + \theta(b^+ - a)|^0 = |\bar{u}_{\rho,r,0}(x) - f_0(x)|^0.$$

For $x \in W_r^-$, we have

$$\begin{aligned}K^- &:= |(1-\theta)(b^+ - a) + \theta(b^- - a)| - |b^- - a| \\ &= |\bar{u}_{\rho,r,0}(x) - f_0(x)| - |u_0(x) - f_0(x)|.\end{aligned}$$

Moreover,

$$\zeta^- := |(1-\theta)(b^+ - a) + \theta(b^- - a)|^0 = |\bar{u}_{\rho,r,0}(x) - f_0(x)|^0.$$

Writing $h_r = h_r^{(+)} - h_r^{(-)}$ for $h_r^{(+)}, h_r^{(-)} \ge 0$, thus

$$\begin{aligned}L &:= \int_{U_r} \phi(\bar{u}_{\rho,r,0}(x) - f_0(x)) - \phi(u_0(x) - f_0(x))\,dx \\ &\le C_\phi K^+ \zeta^+ \int_{V_\Gamma} \rho h_r^{(+)}(v)\,dv + C_\phi K^- \zeta^- \int_{V_\Gamma} \rho h_r^{(-)}(v)\,dv.\end{aligned}$$

As in the proof of Lemma 6.3, we concentrate on the case $b^+ < b^-$ with $a \ge (b^+ + b^-)/2$, setting $\sigma = 1$. Other cases are handled analogously with $\sigma = -1$. Since $b^+ \ne b^-$, there are at most two choices of θ for which $\zeta^+ = 0$ or $\zeta^- = 0$. Simply by the triangle inequality

$$K^- \le (1-\theta)|b^+ - b^-|.$$

Further, choosing $\theta < 1$ large enough while maintaining $\zeta^{\pm} = 1$, we can ascertain

$$K^+ \le (1-\theta)(b^+ - b^-) = -(1-\theta)|b^+ - b^-|.$$

Thus

$$L \le C_\phi |b^+ - b^-| \left(\int_{V_\Gamma} \rho h_r^{(-)}(v)\, dv - \int_{V_\Gamma} \rho h_r^{(+)}(v)\, dv \right) = -C_\phi \rho \lambda_u(x_0) I_r. \qquad \square$$

Remark B.1. Looking at the proof of Lemma 6.3 with a little bit more care in the approximation of K^- above, we could in fact get the stronger double-sided estimate (6.16) if u "jumps through f", in other words, if

$$\min\{u^+(x), u^-(x)\} < \widetilde{f}(x_0) < \max\{u^+(x), u^-(x)\}.$$

B.2. Proof of Lemma 7.2. The proof, as well as the proof of Lemma 7.3, depends on the following technical result on the pointwise R-curvature.

LEMMA B.2. *Let $\alpha > 0$ and $\Gamma \subset \Omega$ be a Lipschitz $(m-1)$-graph. Then at any $x_0 \in \Gamma$, either of the following holds.*

(i) $\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) = \alpha^{-1}$.

(ii) For some $\kappa \in (0,1)$ there exist $r_j \searrow 0$ and for each j, some $\rho_{j,i} \searrow 0$ such that for every $i, j \in \mathbb{Z}^+$ either

$$R(\gamma_{\rho_{j,i}, r_j \#} u) - R(u) - \kappa \alpha^{-1} \rho_{j,i} I_r \le 0, \tag{B.2}$$

or

$$R(\gamma_{\rho_{j,i}, r_j \#} u) - R(u) - \kappa^{-1} \alpha^{-1} \rho_{j,i} I_r \ge 0. \tag{B.3}$$

Proof. We observe that $\gamma_{\rho,r} = \rho\gamma_{1,r} + (1-\rho)\iota$, so that

$$\mathcal{D}_u^R(\gamma_{1,r}) = \limsup_{\rho \searrow 0} \frac{R(\gamma_{\rho,r\#}u) - R(u)}{\rho}.$$

Suppose case (ii) does not hold at $x_0 \in \Gamma$. Fix $\kappa \in (0,1)$. Then there exists $r_0 > 0$ such that for every $0 < r < r_0$, there exists $\rho^r > 0$ such that for every $0 < \rho < \rho^r$ we have bounds

$$\kappa^{-1}\alpha^{-1}\rho I_r \le R(\gamma_{\rho,r\#}u) - R(u) \le \kappa \alpha^{-1} \rho I_r. \tag{B.4}$$

If we define the upper and lower transformation differentials

$$\mathcal{D}_u^{R,*}(\gamma) := \limsup_{\rho \searrow 0} \frac{R(\gamma_\# u) - R(u)}{\rho} \quad \text{and} \quad \mathcal{D}_{u,*}^{R}(\gamma) := \liminf_{\rho \searrow 0} \frac{R(\gamma_\# u) - R(u)}{\rho},$$

then dividing (B.4) by ρ and letting $\rho \searrow 0$, it follows that

$$\kappa^{-1}\alpha^{-1} I_r \le \mathcal{D}_{u,*}^R(\gamma_{\rho,r}) \le \mathcal{D}_u^{R,*}(\gamma_{\rho,r}) \le \kappa\alpha^{-1} I_r.$$

Dividing by I_r and letting $r \searrow 0$, we find

$$\kappa^{-1}\alpha^{-1} \le \liminf_{r \searrow 0} I_r^{-1} \mathcal{D}_{u,*}^R(\gamma_{\rho,r}) \le \limsup_{r \searrow 0} I_r^{-1} \mathcal{D}_u^{R,*}(\gamma_{\rho,r}) \le \kappa \alpha^{-1}$$

Since $\kappa \in (0,1)$ was arbitrary, we deduce

$$\alpha^{-1} = \mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0).$$

In particular, $\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0)$ exists and case (i) holds. $\square$

With this, we are ready to prove the curvature expression claimed in Lemma 7.2.

Proof of Lemma 7.2. By Lemma B.1, given $\epsilon > 0$, we have the existence of $r_0 > 0$ and $\sigma \in \{-1, +1\}$ such that

$$\int_\Omega \phi(\bar{u}_{\rho,r}(x) - f(x))\,dx - \int_\Omega \phi(u(x) - f(x))\,dx \le \epsilon\rho r^m - \sigma(1-\theta)\rho C_\phi \lambda_u(x_0) I_r, \quad \text{(B.5)}$$

whenever $0 < r < r_0$ and $-1 < \rho < 1$. We concentrate on the $\sigma = +1$; the case $\sigma = -1$ is handled analogously (or simply by exchanging sides Γ^+ and Γ^-). We suppose first that Lemma B.2(ii) holds for some $x_0 \in (\Gamma \cap J_u) \setminus (J_f \cup Z_u)$ with $\alpha = 1/(\lambda_u(x_0) C_\phi)$. We further concentrate first on the case that (B.2) holds. Then for some $\kappa \in (0,1)$, a sequence $r_j \searrow 0$, and for each fixed j, a sequence $\rho_{j,i} \searrow 0$, we have

$$R(\bar{u}_{\rho_{j,i},r_j}) - R(u) \le (1-\theta)\big(R(\gamma_{\rho_{j,i},r_j\#}U) - R(u)\big) \le (1-\theta)\kappa C_\phi \rho_{j,i} \lambda_u(x_0) I_{r_j}.$$

Taking $r = r_j$ and $\rho = \rho_{j,i}$ and summing with (B.5), we find that

$$\begin{aligned}\Big(\int_\Omega \phi(\bar{u}_{\rho_{j,i},r_j}(x) - f(x))\,dx + R(\bar{u}_{\rho_{j,i},r_j})\Big) - \Big(\int_\Omega \phi(u(x) - f(x))\,dx + R(u)\Big)& \\ \le \epsilon\rho_{j,i} r_j^m - (1-\kappa)(1-\theta)\rho_{j,i} C_\phi \lambda_u(x_0) I_{r_j},& \\ \le \epsilon\rho_{j,i} r_j^m - C''\rho_{j,i} r_j^m.&\end{aligned}$$

Here the constant $C'' > 0$. Choosing j large enough that we can take $\epsilon = C''/2$, we can make this negative. This yields a contradiction to u solving (P).

Suppose then that (B.3) holds. Then for some $\kappa > 0$ we have

$$R(\bar{u}_{\rho_{j,i},r_j}) - R(u) \ge (1-\theta)\kappa^{-1} C_\phi \rho_{j,i} \lambda_u(x_0) I_{r_j}, \quad \text{(B.6)}$$

We now use the double-Lipschitz comparability we already used in the proof of Lemma 6.5. Namely, choosing $r > 0$ small enough, we have for some constant $C' > 0$ that

$$R(\bar{u}_{\rho,r}) + R(\bar{u}_{-\rho,r}) - R(u) \le C'\rho^2 r^{m-1}.$$

We may w.log assume that all $\{r_i\}_{i \in \mathbb{Z}^+}$ are small enough for this. Using (B.6) it follows

$$R(\bar{u}_{-\rho_{j,i},r_j}) - R(u) \le C'\rho_{j,i}^2 r_j^{m-1} - (1-\theta)\kappa^{-1} C_\phi \rho_{j,i} \lambda_u(x_0) I_{r_j}. \quad \text{(B.7)}$$

Summing (B.5) with (B.7) we find that

$$\begin{aligned}\Big(\int_\Omega \phi(\bar{u}_{-\rho_{j,i},r_j}(x) - f(x))\,dx + R(\bar{u}_{-\rho_{j,i},r_j})\Big) - \Big(\int_\Omega \phi(u(x) - f(x))\,dx + R(u)\Big)& \\ \le C'\rho_{j,i}^2 r_j^{m-1} + \epsilon\rho_{j,i} r_j^m + (1-\kappa^{-1})(1-\theta)\rho_{j,i} C_\phi \lambda_u(x_0) I_{r_j}& \\ \le C'\rho_{j,i}^2 r_j^{m-1} + \epsilon\rho_{j,i} r_j^m - C''\rho_{j,i} r_j^m.&\end{aligned}$$

Again the constant $C'' > 0$. Choosing $j \in \mathbb{Z}^+$ large enough that we can take $0 < \epsilon < C''/2$, and then choosing i large enough that $\rho_{j,i}^2$ becomes very small, we can make the right hand side negative. Thus we again have a contradiction to u solving (P). It follows that Lemma B.2(i) has to hold for all $x_0 \in (\Gamma \cap J_u) \setminus (J_f \cup Z_u)$. With our choice $\alpha = 1/(\lambda_u(x_0) C_\phi)$, this says exactly that $\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) = \lambda_u(x_0) C_\phi$ for $\mathcal{H}^{m-1}$-a.e. $x_0 \in (J_u \setminus J_f) \cap \Gamma$. □

B.3. Proof of Lemma 7.3. We use the next lemma to translate the pointwise R-curvature for general R into pointwise TV_Γ-curvature for $\mathrm{TV}_\Gamma(u) := |Du|(\Gamma)$ under the approximate piecewise constancy assumption.

LEMMA B.3. *Let $\Gamma \subset \Omega$ be a Lipschitz $(m-1)$-graph and $x_0 \in \Gamma$. Suppose R is separably double-Lipschitz comparable, and $\Theta_m(|Du| \llcorner \Omega \setminus \Gamma; x_0) = 0$. Then*

$$\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) = \lim_{r \searrow 0} I_r^{-1} \widetilde{\mathcal{D}}_u^R(\gamma_{1,r}), \tag{B.8}$$

where

$$\widetilde{\mathcal{D}}_u^R(\gamma_{1,r}) := \lim_{\rho \searrow 0} R^s \frac{|D\gamma_{\rho,r\#}u|(\gamma_{\rho,r}(\Gamma)) - |Du|(\Gamma)}{\rho}.$$

Proof. We recall that

$$\mathcal{D}_u^R(\gamma_{1,r}) = \lim_{\rho \searrow 0} \frac{R(\gamma_{\rho,r\#}u) - R(u)}{\rho}.$$

By Definition 3.3 we have

$$R(\gamma_{\rho,r\#}u) - R(u) \le R^a T_{\gamma_{\rho,r},\iota} |Du|(\operatorname{cl} U \setminus \Gamma) + R^s \big(|D\gamma_{\rho,r\#}u|(\gamma_{\rho,r}(\Gamma)) - |Du|(\Gamma)\big).$$

Using Lemma 5.4 to estimate $T_{\gamma_{\rho,r},\iota}$, and minding the assumption $\Theta_m(|Du| \llcorner \Omega \setminus \Gamma; x_0) = 0$, we deduce for arbitrary $\epsilon > 0$, for small enough $r > 0$ that

$$\mathcal{D}_u^R(\gamma_{1,r}) \le \epsilon r^m + R^s \limsup_{\rho \searrow 0} \frac{|D\gamma_{\rho,r\#}u|(\gamma_{\rho,r}(\Gamma)) - |Du|(\Gamma)}{\rho}.$$

Dividing by I_r, letting $r \searrow 0$, and afterwards letting $\epsilon \searrow 0$, yields the upper bound

$$\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) \le \lim_{r \searrow 0} I_r^{-1} \widetilde{\mathcal{D}}_u^R(\gamma_{1,r}).$$

In order to derive the corresponding lower bound, we set $\gamma := \gamma_{\rho,r}$ and $v := \gamma_\# u$. Then

$$\begin{aligned} R(\gamma_\# u) - R(u) = R(v) - R({\gamma^{-1}}_\# v) &\ge -R^a T_{\gamma^{-1},\iota} |Du|(\operatorname{cl} U \setminus \gamma(\Gamma)) \\ &\quad - R^s \frac{|D\gamma_\# v|(\Gamma) - |Dv|(\gamma(\Gamma))}{\rho}. \end{aligned}$$

Using Lemma 5.4 to estimate $T_{\gamma^{-1},\iota}$, and proceeding as above, we deduce (B.8). ☐

We now proceed with the final missing proof.

Proof of Lemma 7.3. We define the shorthand notation $\lambda(v) := \lambda_u(g_\Gamma(v))$, and observe from Lemma A.1 that

$$|D\gamma_{\rho,r\#}u|(\gamma_{\rho,r}(\Gamma)) = \int_{V_\Gamma} \lambda(v) \sqrt{1 + \|\nabla(f_\Gamma + \rho h_r)(v)\|^2}\, dv,$$

and

$$|Du|(\Gamma) = \int_{V_\Gamma} \lambda(v) \sqrt{1 + \|\nabla f_\Gamma(v)\|^2}\, dv.$$

We further calculate using convexity of the norm, and concavity of the square root that

$$
\begin{aligned}
\rho c_{\lambda,f}(h_r) \le \int_V \lambda(v)\sqrt{1+\|\nabla(f_\Gamma+\rho h_r)(v)\|^2}\,dv - \int_V \lambda(v)\sqrt{1+\|\nabla f_\Gamma(v)\|^2}\,dv \\
\le \rho c_{\lambda,f}(h_r) + \rho^2 \int_V \lambda(v) \frac{\|\nabla h_r(v)\|^2}{\sqrt{1+\|\nabla f_\Gamma(v)\|^2}}\,dv,
\end{aligned}
$$

where

$$
c_{\lambda,f}(h) := \int_V \lambda(v) \left\langle \frac{\nabla f_\Gamma(v)}{\sqrt{1+\|\nabla f_\Gamma(v)\|^2}}, \nabla h(v) \right\rangle dv.
$$

Thus

$$
\widetilde{\mathcal{D}}_u^R(\gamma_{1,r}) = R^s c_{\lambda,f}(h_r).
$$

Minding Lemma B.3, we deduce

$$
\mathcal{C}_u^{R,\Gamma,\bar{h}}(x_0) = \lim_{r\searrow 0} R^s c_{\lambda,f}(h_r/I_r).
$$

Since ∇f_Γ is bounded, the claim is immediate if $g_\Gamma^{-1}(x_0)$ is a Lebesgue point of λ and ∇f_Γ, that is $\mathcal{H}^{m-1}$-a.e. □

REFERENCES

[1] W.K. Allard, *Total variation regularization for image denoising, I. Geometric theory*, SIAM Journal on Mathematical Analysis, 39 (2008), pp. 1150–1190.

[2] F. Alter, V. Caselles, and A. Chambolle, *A characterization of convex calibrable sets in* $\mathbb{R}^n$, Mathematische Annalen, 332 (2005), pp. 329–366.

[3] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, 2000.

[4] M. Benning and M. Burger, *Ground states and singular vectors of convex variational regularization methods*, Methods and Applications of Analysis, 20 (2013), pp. 295–334.

[5] A. L. Bertozzi and J. B. Greer, *Low-curvature image simplifiers: Global regularity of smooth solutions and Laplacian limiting schemes*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 764–790.

[6] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences, 3 (2011), pp. 492–526.

[7] K. Bredies, K. Kunisch, and T. Valkonen, *Properties of L^1-TGV^2: The one-dimensional case*, Journal of Mathematical Analysis and Applications, 398 (2013), pp. 438–454.

[8] M. Burger, M. Franek, and C.-B. Schönlieb, *Regularized regression and density estimation based on optimal transport*, Applied Mathematics Research eXpress, 2012 (2012), pp. 209–253.

[9] V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Modeling and Simulation, 6 (2008), pp. 879–894.

[10] A. Chambolle, *An algorithm for mean curvature motion*, Interfaces and Free Boundaries, 6 (2004), p. 195.

[11] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numerische Mathematik, 76 (1997), pp. 167–188.

[12] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM Journal on Scientific Computation, 22 (2000), pp. 503–516.

[13] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM Journal on Applied Mathematics, 65 (2005), pp. 1817–1837.

[14] T. F. Chan, S. H. Kang, and J. Shen, *Euler's elastica and curvature-based inpainting*, SIAM Journal on Applied Mathematics, (2002), pp. 564–592.

[15] B. Dacorogna and J. Moser, *On a partial differential equation involving the jacobian determinant*, Annales de l'Institut Henri Poincaré. Analyse non linéaire, 7 (1990), pp. 1–26.
[16] G. Dal Maso, I. Fonseca, G. Leoni, and M. Morini, *A higher order model for image restoration: the one-dimensional case*, SIAM Journal on Mathematical Analysis, 40 (2009), pp. 2351–2391.
[17] S. Didas, J. Weickert, and B. Burgeth, *Properties of higher order nonlinear diffusion filtering*, Journal of Mathematical Imaging and Vision, 35 (2009), pp. 208–226.
[18] V. Duval, J. F. Aujol, and Y. Gousseau, *The TVL1 model: A geometric point of view*, Multiscale Modeling and Simulation, 8 (2009), pp. 154–189.
[19] I. Fonseca and G. Leoni, *Modern methods in the calculus of variations: L^p spaces*, Springer Verlag, 2007.
[20] D. Gilbarg and N.S. Trudinger, *Elliptic partial differential equations of second order*, Classics in mathematics, Springer, 2001.
[21] P. Guidotti, *Anisotropic diffusions of image processing from Perona-Malik on*, Advanced Studies in Pure Mathematics, (2014). To appear.
[22] M. Hintermüller and G. Stadler, *An infeasible primal-dual algorithm for total bounded variation–based inf-convolution-type image restoration*, SIAM Journal on Scientific Computation, 28 (2006), pp. 1–23.
[23] M. Hintermüller and T. Wu, *Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 1385–1415.
[24] ———, *A superlinearly convergent R-regularized Newton scheme for variational models with concave sparsity-promoting priors*, Computational Optimization and Applications, 57 (2014), pp. 1–25.
[25] M. Hintermüller, T. Valkonen, and T. Wu, *Limiting aspects of non-convex TV^φ models.* Submitted, 2014.
[26] J. Huang and D. Mumford, *Statistics of natural images and models*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, 1999.
[27] J. Lellmann, D. Lorenz, C.-B. Schönlieb, and T. Valkonen, *Imaging with Kantorovich-Rubinstein discrepancy*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 2833–2859.
[28] M. Lysaker, A. Lundervold, and X.-C. Tai, *Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time*, IEEE Transactions on Image Processing, 12 (2003), pp. 1579–1590.
[29] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations*, American Mathematical Society, 2001.
[30] M. Nikolova, *Minimizers of cost functions involving nonsmooth data-fideltiy terms. Application of processing of outliers*, SIAM Journal on Numerical Analysis, 40 (2002), pp. 965–994.
[31] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, *An iterated l1 algorithm for non-smooth non-convex optimization in computer vision*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.
[32] K. Papafitsoros and K. Bredies, *A study of the one dimensional total generalised variation regularisation problem.* Preprint, 2013.
[33] K. Papafitsoros and C.-B. Schönlieb, *A combined first and second order variational approach for image reconstruction*, Journal of Mathematical Imaging and Vision, 48 (2014), pp. 308–338.
[34] K. Papafitsoros and T. Valkonen, *Asymptotic behaviour of total generalised variation*, in Fifth International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), 2015. Accepted.
[35] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (1990), pp. 629–639.
[36] F. Rindler and G. Shaw, *Strictly continuous extensions of functionals with linear growth to the space BV.* preprint, 2013.
[37] W. Ring, *Structural properties of solutions to total variation regularization problems*, ESAIM: Mathematical Modelling and Numerical Analysis, 34 (2000), pp. 799–810.
[38] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[39] J. Shen, S. Kang, and T. Chan, *Euler's elastica and curvature-based inpainting*, SIAM Journal on Applied Mathematics, 63 (2003), pp. 564–592.
[40] T. Valkonen, *Transport equation and image interpolation with SBD velocity fields*, Journal de mathématiques pures et appliquées, 95 (2011), pp. 459–494.
[41] ———, *The jump set under geometric regularisation. Part 2: Higher-order approaches.* Submitted, July 2014.

[42] ———, *A primal-dual hybrid gradient method for non-linear operators with applications to MRI*, Inverse Problems, 30 (2014), p. 055012.
[43] L. A. Vese and S. J. Osher, *Modeling textures with total variation minimization and oscillating patterns in image processing*, Journal of Scientific Computing, 19 (2003), pp. 553–572.
[44] J. Weickert, *Anisotropic Diffusion in Image Processing*, ECMI series, B.G. Teubner Stuttgart, 1998.
[45] W. Yin, D. Goldfarb, and S. Osher, *The total variation regularized L^1 model for multiscale decomposition*, Multiscale Modeling and Simulation, 6 (2007), pp. 190–211.