The jump set under geometric regularisation. 
Part 1: Basic technique and first-order denoising
Tuomo Valkonen*

Abstract
Let $u \in BV(\Omega)$ 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 $H^{m-1}(J_u \setminus J_f) = 0$ 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.

Mathematics subject classification: 26B30, 49Q20, 65J20.

Keywords: total variation, jump set, regularisation, Lipschitz, Huber, Perona-Malik.

1. Introduction
We study the structure of the approximate jump set $J_u$ of solutions $u \in BV(\Omega)$ 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 $H^{m-1}$-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 $u \in BV(\Omega)$ and that $R$ behaves almost like TV under small Lipschitz transformations. We state these conditions in detail in Section 3. Briefly, we require the BV-coercivity condition

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

and a double-Lipschitz comparability condition of the form

$$R(\gamma \cdot u) + R(\overline{\gamma \cdot u}) - 2R(u) \leq CT_{\gamma,\overline{\gamma}}|Du|(clU). \tag{1.1}$$

Here $\gamma$ and $\overline{\gamma}$ are Lipschitz transformations on an open set $U$. The suitably defined distance $T_{\gamma,\overline{\gamma}}$ between the transformations turns out to be $O(\rho^2)$ for specially constructed shift transformations, dependent on a parameter $\rho$. As we will see in Section 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 at the extension of this condition and technique to cover higher-order regularisers, in particular total generalised variation (TGV) [7], and infimal convolution TV (ICTV) [12]. 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.

* Prometeo Fellow, Research Center on Mathematical Modeling (Modemat), Escuela Politécnica Nacional de Quito, Ecuador. E-mail: tuomo.valkonen@iki.fi.
Assuming further that both the solution $u$ and the given data $f$ are in $\text{BV}(\Omega) \cap L^\infty_{\text{loc}}(\Omega)$, we show for $p$-increasing $\phi$ for $1 < p < \infty$, including $L^p$ fidelities $\phi(t) = |t|^p$, that the jump set $J_u$ of $u$ is contained, modulo a $H^{m-1}$ null set, in the jump set $J_f$ of $f$. That is, $H^{m-1}(J_u \setminus J_f) = 0$. 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 \[19\]. 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 $\alpha$ the (asymptotic) regularisation parameter. We state all of these results in detail in Section \[4\]. The proofs are split between multiple sections. We construct the specific Lipschitz transformations in Section \[3\]. The main part of the general proof, studying the effect of the transformations on the fidelity, can be found in Section \[3\] for $p > 1$, and in Section \[4\] for $p = 1$.

The class of problems \[3\] is of importance, in particular, for image denoising. From 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) \[3\] and infimal convolution TV \[3\], 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 \[3\], with varying fidelity $\phi$ 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 \[26\], and besides TGV and ICTV, various other higher-order regularisation schemes have been proposed in the recent years \[4, 27, 13, 11, 13\]. Curvature based regularisers such as Euler’s elastica \[26, 35\] and \[11\] 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 \[23, 24, 22, 32, 35\]. In the other direction, “lower-order schemes” such as Meyer’s G-norm \[36, 35\] and TV with Kantorovich-Rubinstein discrepancy \[28\] 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 Section \[4\]. We also demonstrate that our technique would apply to non-convex TV schemes, as well as the Perona-Malik anisotropic diffusion \[13, 14\], if these were well-posed to begin with \[35, 22\]. 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 \[4\] on one-dimensional domains $\Omega = (a, b)$ \[11, 35\]. In particular, we showed in \[1\] 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. $R(u) = \alpha TV(u) = \alpha \|Du\|_{M(\Omega; \mathbb{R}^m)}$, the literature is more plentiful on analytical results. With the squared $L^2$ fidelity, the problem \[7\] is also known as the Rudin-Osher-Fatemi \[35\] problem (ROF), and written
\[
\min_{u \in \text{BV}(\Omega)} \frac{1}{2} \int_{\Omega} |f(x) - u(x)|^2 \, dx + \alpha TV(u).
\]

The first structural results can be found in \[17\], where the stair-casing property is studied. Regarding the jump set, it is shown in \[17\] that $H^{m-1}(J_u \setminus J_f) = 0$. 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
\[
TV(u) = \int_{-\infty}^{\infty} \text{Per}(\{u > t\}; \Omega) \, dt,
\]

where $\text{Per}(E; \Omega) := \|DX\|_{L^2(\Omega; \mathbb{R}^m)}$ denotes the perimeter of the set $E$ within $\Omega$. Thanks to the co-area formula, the problem \[12\] can be shown to be equivalent to the family of minimal surface problems
\[
\min_{E \subseteq \Omega} \int_{E} \text{sgn}(t)(f(x) - u(x)) \, dx + \alpha \text{Per}(E; \Omega), \quad (\text{a.e. } t \in \mathbb{R}).
\]

The level sets of $u$ can be found as solutions of \[4\] for varying $t$. This formulation and the regularity of the level sets is studied in \[3, 11, 14\].

For TV regularisation with the $L^1$ fidelity $\phi(x) = x$, i.e., the problem
\[
\min_{u \in L^1(\Omega)} \int_{\Omega} |f(x) - u(x)| \, dx + \alpha TV(u),
\]

\[15\]
various basic properties were first studied by [4]. These include thresholds for the regularisation parameter \( \alpha > 0 \), under which the optimal solution \( u \) equals \( f \). Similar thresholds were also derived in [3] 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 [5]. In [6] it is shown that the level sets \( E_c = \{ u \geq c \} \) of solutions \( u \) to (1.5), are solutions to the minimal surface problems

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

Some further properties of the level sets are studied in [7]. 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 \( \alpha^{-1} \) for \( \alpha \) the regularisation parameter. In a similar fashion, the capability of (1.5) to separate geometric objects according to their scales is studied in [8]. Finally, an interesting direction is taken in [9] by studying singular vectors and ground states of regularisation functionals, eventually hoping to obtain something resembling an eigendecomposition for solutions to (7). 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 (3) if \( \mathcal{H}^{m-1}(J_u \setminus J_f) > 0 \), \( (p > 1) \). Before starting to develop this technique in detail in Section 4, we now introduce some general notation, concepts and tools in the following Section 4.

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 \( \partial A \), and the closure by \( \text{cl} A \). The \( \{0,1\} \)-valued indicator function we write as \( \chi_A \). We denote the open ball of radius \( \rho \) centred at \( x \in \mathbb{R}^m \) by \( B(x, \rho) \). We denote by \( \omega_m \) 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 \( \nu \), 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 \( r(A) \) 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 \( \Omega \) by \( \mathcal{M}(\Omega) \). If \( V \) is a vector space, then the space of Radon measures on \( \Omega \) with values in \( V \) is denoted \( \mathcal{M}(\Omega; V) \). The \( k \)-dimensional Hausdorff measure, on any given ambient space \( \mathbb{R}^m \), \( (k \leq m) \), is denoted by \( \mathcal{H}^k \), while \( L^m \) denotes the Lebesgue measure on \( \mathbb{R}^m \). The total variation (Radon) norm of a measure \( \mu \) is denoted \( ||\mu||_{\mathcal{M}(\mathbb{R}^m)} \).

For vector-valued measures \( \mu \in \mathcal{M}(\Omega; \mathbb{R}^k) \), we use the notation \( ||\mu||_{q,\mathcal{M}(\mathbb{R}^m)} \) to indicate that the finite-dimensional base norm is the \( q \)-norm. We use the same notation for vector fields \( w \in L^p(\Omega; \mathbb{R}^k) \), namely

\[
||w||_{q,L^p(\Omega)} := \left( \int_\Omega ||w(x)||_q^p \, dx \right)^{1/p}.
\]

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

We denote the \( k \)-dimensional upper resp. lower density of \( \mu \) 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 \( L(V; W) \) the space of linear maps between the vector spaces \( V \) and \( W \). If \( L \in L(V; \mathbb{R}^k) \), where \( V \sim \mathbb{R}^n \), \((n \leq k)\), is a finite-dimensional Hilbert space, then \( L^* \in L(\mathbb{R}^k; V^*) \) denotes the adjoint, and the \( n \)-dimensional Jacobian is defined as

\[
J_n[L] := \sqrt{\det(L^* \circ L)}.
\]

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

\[
J_n[f](x) := J_n[(\nabla f(x))^*].
\]

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

\[
\int_{\mathbb{R}^k} \sum_{x \in \Omega \cap f^{-1}(y)} g(x) dH^n(y) = \int_{\Omega} g(x)J_n[f](x) dH^n(x).
\] (2.1)

That this indeed holds in our sitting 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 H"{o}lder continuous second differential for all exponents \( \lambda \in (0,1) \).

The Lipschitz factor of a Lipschitz mapping \( f \) we denote by \( \text{lip} f \). 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 1-to-1 Lipschitz transformation, and \( u : \Omega \to \Omega \) a Borel function, we define the pushforward \( u_* := \gamma \# u := u \circ \gamma^{-1} \). 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 \( z_\Gamma \), 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” \( \Gamma \), 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 \( f_\Gamma \) and \( g_\Gamma \) from \( V_\Gamma \) to the whole space \( z_\Gamma^\perp \) without altering the Lipschitz constant. Then \( \Gamma \) splits \( \Omega \) 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 $u : \Omega \rightarrow \mathbb{R}$ on a bounded open set $\Omega \subset \mathbb{R}^m$ is of bounded variation (see, e.g., Proposition 3.92) for a more thorough introduction), denoted $u \in BV(\Omega)$, if $u \in L^1(\Omega)$, and the distributional gradient $Du$ is a Radon measure. Given a sequence $\{u^i\}_{i=1}^\infty \subset BV(\Omega)$, weak* convergence is defined as $u^i \rightharpoonup u$ strongly in $L^1(\Omega)$ along with $Du^i \rightharpoonup Du$ weakly* in $\mathcal{M}(\Omega)$. The sequence converges strictly if, in addition to this, $|Du^i|(\Omega) \rightarrow |Du|(\Omega)$.

We denote by $S_u$ the approximate discontinuity set, i.e., the complement of the set where the Lebesgue limit $\bar{u}$ exists. The latter is defined by
\[
\lim_{\rho \searrow 0} \frac{1}{\rho^m} \int_{B(x, \rho)} \|\bar{u}(x) - u(y)\| \, dy = 0.
\]

The distributional gradient can be decomposed as $Du = \nabla u \mathcal{L}^m + D^j u + D^s u$, where the density $\nabla u$ 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
\[
\lim_{\rho \searrow 0} \frac{1}{\rho^m} \int_{B(x, \rho) \setminus B^\pm(x, \rho, \nu)} \|u^\pm(x) - u(y)\| \, dy = 0,
\]
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 $\nu$ 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 $\sigma$-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., Proposition 3.92.

**Proposition 2.1.** Let $u \in BV(\Omega)$ for an open domain $\Omega \subset \mathbb{R}^m$. Define
\[
\tilde{S} := \{x \in \Omega \mid \Theta_{*, m}(|Du|; x) = \infty\}, \quad \text{and} \quad \tilde{J} := \{x \in \Omega \mid \Theta_{*, m-2}(|Du|; x) > 0\}.
\]
Then the following decomposition holds.
\begin{enumerate}[(i)]
\item $\nabla u = Du_\llcorner (\Omega \setminus \tilde{S})$.
\item $D^j u = Du_\llcorner \tilde{J}$, precisely $\tilde{J} \supset J_u$, and $\mathcal{H}^{m-1}(\tilde{J} \setminus J_u) = 0$.
\item $D^c u = Du_\llcorner (\tilde{S} \setminus J)$.
\end{enumerate}

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 ($p > 1$) $\phi$. We just have to rigorously state our assumptions on the regularisation functional $R$ and the fidelity $\phi$. These ensure firstly the existence of a solution $u \in BV(\Omega)$ to (3.1). Secondly we want to ensure that $R$ behaves almost like TV under averaged Lipschitz transformations, and $\phi$ 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 $BV(\Omega)$, and there exists $C > 0$ such that
\[
\|Du\|_{2, \mathcal{M}(\Omega; \mathbb{R}^m)} \leq C(1 + \|u\|_{L^1(\Omega)} + R(u)), \quad (u \in L^1(\Omega)).
\]
**Definition 3.2.** We denote by $\mathcal{F}(\Omega)$ the set of one-to-one Lipschitz transformations $\gamma : \Omega \to \Omega$ with $\gamma^{-1}$ 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 $\gamma, \gamma' \in \mathcal{F}(\Omega)$, we then define the basic double-Lipschitz comparison constants

$$
G_{\gamma, \gamma'} := \sup_{x \in \Omega, \|v\| = 1} \|A_{\gamma'}(x)v\| + \|A_{\gamma}(x)v\| - 2\|v\|,
$$

and

$$
J_{\gamma, \gamma'} := \sup_{x \in \Omega} |J_m \gamma(x) + J_m \gamma'(x) - 2|.
$$

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

$$
A_{\gamma}(x) := \nabla \gamma^{-1}(\gamma(x))J_m \gamma(x).
$$

We also define the distance-to-identity

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

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

$$
T_{\gamma, \gamma'} := G_{\gamma, \gamma'} + J_{\gamma, \gamma'} + D_{\gamma}^2 + D_{\gamma'}^2.
$$

**Remark 3.1.** With the help of $J_{\gamma, \gamma'}$, we could generally replace $G_{\gamma, \gamma'}$ by

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

It however turns out that for our transformations of interest, $A_{\gamma}$ is easier work with than $\nabla \gamma^{-1} \circ \gamma$ directly.

**Definition 3.3.** We say that $R$ is double-Lipschitz comparable if there exists a constant $R^a = R^a(\Omega)$ such that for every $u \in BV(\Omega)$, every open set $U \subset \Omega$, and every $\gamma, \gamma' \in \mathcal{F}(\Omega, U)$ with $T_{\gamma, \gamma'} < 1$ we have

$$
R(\gamma\#u) + R(\gamma'\#u) - 2R(u) \leq R^a T_{\gamma, \gamma'}|Du|(cl U).
$$

We also say that $R$ is separably double-Lipschitz comparable if there exist constants $R^a = R^a(\Omega)$ and $R^2 = R^2(\Omega)$ such that for every $u \in BV(\Omega)$, every open set $U \subset \Omega$, and every $\gamma, \gamma' \in \mathcal{F}(\Omega, U)$ with $T_{\gamma, \gamma'} < 1$ we have

$$
R(\gamma\#u) + R(\gamma'\#u) - 2R(u) \leq R^a T_{\gamma, \gamma'}|Du|(cl U \setminus \Gamma)
$$

$$
+ R^2 (|D^\gamma\#u|_{(\gamma'(\Gamma))}) + |D^\gamma\#u|_{(\gamma(\Gamma))} - 2|Du|_{(\Gamma)}), \quad (\gamma, \gamma' \in \mathcal{F}(\Omega, U)).
$$

**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 $T_{\gamma, \gamma'} < 1$ is arbitrary; we only need some bound for Huber-regularised TV.

In order to show the existence of solutions to (3.2), we require the following property from $\phi$.

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

$$
(\|u\|_{L^1(\Omega)}) \leq C \left( \int_{\Omega} \phi(|u(x)|) \, dx + 1 \right), \quad (u \in L^1(\Omega)). \tag{3.2}
$$

Throughout this paper, we extend the domain of $\phi$ to $\mathbb{R}$ by defining

$$
\phi(t) := \phi(-t), \quad (t < 0).
$$

This is in order to simplify the notation $\phi(|u(x)|)$ to $\phi(u(x))$.

For the study of the jump set $J_u$ of solutions to (3.2), we require additionally the following increase criterion to be satisfied by $\phi$. 

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

$$\phi(x) - \phi(y) \leq C_\phi (x - y)|x|^{p-1}, \quad (x, y \geq 0).$$

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

Example 3.1. Let $\phi(x) = x^p$, $(p \geq 1)$. Then $\phi$ 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 \([\text{I}]\) is well-posed under the above assumptions. We may therefore proceed with the analysis of the structure of the solutions $u \in BV(\Omega)$.

Theorem 3.1. 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 $\phi$ an admissible fidelity function for $\Omega$. Then there exists a solution $u \in L^1(\Omega)$ to \([\text{I}]\), and any solution satisfies $u \in BV(\Omega)$.

Proof. Clearly the minimum in \([\text{I}]\) is finite. Let $\{u^i\}_{i=0}^\infty$ be a minimising sequence for \([\text{I}]\). Minding that \([\text{I}]\), \([\text{II}]\), bound $\{u^i\}_{i=0}^\infty$ in $BV(\Omega)$, and that $R$ is lower semi-continuous with respect to weak$^*$ convergence in $BV(\Omega)$, and $\phi$ is lower semi-continuous, the claim follows by the standard method of calculus of variations, see, e.g., [4].

3.2. The main results

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

Theorem 3.2. 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 $f \in BV(\Omega) \cap L^\infty(\Omega)$, and suppose $u \in BV(\Omega) \cap L^\infty(\Omega)$ solves \([\text{I}]\). Then

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

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

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

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

$$C_{R,\Lambda_i}(x) = |u^+(x) - u^-(x)|C_\phi.$$  \hspace{1cm} (3.3)

This technical definition will be provided later in Definition 7.4.

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

$$-\nabla \frac{\nabla f_{\Lambda_i}(v)}{\sqrt{1 + \|\nabla f_{\Lambda_i}(v)\|^2}} = C_\phi/R^s, \quad (v \in V_{\Lambda_i}).$$  \hspace{1cm} (3.4)

Remark 3.5. It is not very difficult to improve Theorem 3.3 a little bit. Namely, we can replace the assumption $\Theta_m(|Du|; \Omega \setminus \Lambda_i; x) = 0$ by that of $\nabla u$ having one-sided Lebesgue limits at $x$, with corresponding normal $v = \nu_{J_u}(x)$. 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 3.3 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.
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 we will concentrate on higher-order regularisation functionals: second-order total generalised variation (TGV²), and infimal convolution TV (ICTV), whose analysis is more involved. We however remark that both Theorem 4.1 and Theorem 4.3 can easily be derived for ICTV from the corresponding result for TV. For TGV this is not the case.

4.1. Total variation

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

\[
\text{TV}(u) := |Du|(\Omega) = \sup \left\{ \int_{\Omega} (\text{div} \varphi(x), u(x)) \, dx : \varphi \in C^\infty_c(\Omega), \|\varphi\|_{L^\infty(\Omega)} \leq 1 \right\}.
\]

We now show that it satisfies the conditions of Theorem 3.2.

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

We use the following simple lemma for the proof.

**Lemma 4.1.** Let \( u \in \text{BV}(\Omega) \) 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 \( \gamma(x) = \gamma(\gamma^{-1}(x)) \) satisfies \( \nabla \gamma(x) = \nabla \gamma(\gamma^{-1}(x)) \nabla \gamma^{-1}(x) \) for \( \mathcal{L}^m \)-a.e. \( x \in \Omega \). Repeating the same argument on the formulation \( \gamma(x) = \gamma^{-1}(\gamma(x)) \), therefore \( \nabla \gamma(\gamma^{-1}(x)) \nabla \gamma^{-1}(x) = \nabla \gamma^{-1}(\gamma(x)) \), 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 3.2 are trivial in this case. Also \( \|u\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)} \) is well-known for solutions \( u \) to (F). 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 \( u \in \text{BV}(\Omega) \). By Lemma 3.1, we have

\[
|D\gamma \# u|(\Omega) = |D^a \gamma \# u|(\Omega) + |D^s \gamma \# u|(\Omega) = |\nabla \gamma^{-1} \gamma \# \nabla u \mathcal{L}^m|(\Omega) + |D^s \gamma \# u|(\Omega)
\]

(4.1)

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

\[
|\nabla \gamma^{-1} \gamma \# \nabla u \mathcal{L}^m|(\Omega) = \int_{\Omega} \|\nabla \gamma^{-1}(x) \nabla u(\gamma^{-1}(x))\| \, dx
\]

\[
= |D^a u|(\Omega \setminus U) + \int_U \|A_\gamma(x) \nabla u(x)\| \, dx.
\]

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

\[
|\nabla \gamma^{-1} \gamma \# \nabla u \mathcal{L}^m|(\Omega) + |\nabla \gamma^{-1} U \# \nabla u \mathcal{L}^m|(\Omega) - 2|D^a u|(\Omega) \leq G_{\gamma, U} |D^a u|(\Omega).
\]
Recalling (L.1), and minding that
\[ |D^s\gamma#u|(\Omega \setminus U) = |D^s\gamma#u|\Omega \setminus U) = |D^s\gamma#u|\Omega \setminus U), \]
we deduce
\[ |D^s\gamma#u|(\Omega) + |D\gamma#u|\Omega) - 2|Du|\Omega) \leq G_{\gamma,\gamma}[D^s\gamma#u]U) + (|D^s\gamma#u|U) + |D^s\gamma#u|U) - 2|D^s\gamma#u|U) \]. (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 sequence \( \{u^i\}_{i=1}^\infty \subset C^1(\Omega) \), converging strictly in BV(\Omega) to \( u \). Then \( D^s u^i = 0 \), so (L.2) proves
\[ |D^s\gamma#u^i|(\Omega) + |D\gamma#u^i|\Omega) - 2|Du^i|\Omega) \leq G_{\gamma,\gamma}[D^s\gamma#u^i]U) \quad (i = 1, 2, 3, \ldots). \]
Since the strict convergence of \( \{u^i\}_{i=1}^\infty \) to \( u \) bounds the right hand side, we deduce
\[ \sup_i |D\gamma#u^i|\Omega) + \sup_i |D\gamma#u^i|\Omega) < \infty. \]
We may therefore extract a subsequence, unlabelled, such that both \( \{\gamma#u^i\}_{i=1}^\infty \) and \( \{\gamma#u^i\}_{i=1}^\infty \) are convergent weakly* to some \( \bar{u} \in BV(\Omega) \) and \( u \in BV(\Omega) \), respectively. Moreover, by lower semicontinuity of the total variation, and strict convergence of the approximating sequence
\[ |D\gamma|\Omega) + |D\gamma|\Omega) - 2|Du|\Omega) \leq \liminf_i |D\gamma|\Omega) + |D\gamma|\Omega) - 2|Du|\Omega) \]
\[ \leq \liminf_i G_{\gamma,\gamma}[Du^i]U). \]
Let us pick an open set \( U' \supset U \) such that \( |Du|\partial U' = 0 \). Then \( |Du^i|\Omega) \to |Du^i|\Omega \) because \( u^i \to u \) strictly in \( L^1(\Omega) \); see [L. Proposition 1.62]. It follows
\[ |D\gamma|\Omega) + |D\gamma|\Omega) - 2|Du|\Omega) \leq G_{\gamma,\gamma}[Du]U). \]
By taking the intersection over all admissible \( U' \supset U \), we deduce
\[ |D\gamma|\Omega) + |D\gamma|\Omega) - 2|Du|\Omega) \leq G_{\gamma,\gamma}[Du]U). \]
This is almost the double-Lipschitz comparability. We just have to show that \( \bar{u} = \gamma#u \) and \( u = \gamma#u \). Indeed
\[ \int_\Omega |\bar{u}(x) - \gamma#u(x)| \leq \int_\Omega |\bar{u}(x) - \gamma#u^i(x)| \leq \int_\Omega |\gamma#u(x) - \gamma#u^i(x)| \leq \int_\Omega |\bar{u}(x) - \gamma#u^i(x)| \leq \int_\Omega |u(x) - u^i(x)| \leq \int_\Omega |u(x) - u^i(x)|. \]
But \( J_n\gamma(x) \leq (\text{lip } \gamma)^n \infty \), and the integrals on the right hand side tend to zero by the strict convergence of \( u^i \) to \( u \) and the weak* convergence of \( \gamma#u^i \) to \( \bar{u} \). It follows that \( \bar{u} = \gamma#u \). Analogously we show that \( u = \gamma#u \).
Double-Lipschitz comparability is now immediate from (L.3).

Finally, to see separable double-Lipschitz comparability, we proceed analogously as above, but approximate \( u \) on both sides of \( \Omega \). More specifically, referring to Kirszbraun’s theorem, we may assume that \( V^\perp = z^\perp \). Thus \( \Omega \) splits into two domains \( \Omega^\pm = \Omega \cap \Gamma^\pm \). We approximate \( u \) separately on both \( \Omega^+ \) and \( \Omega^- \) by strictly converging sequences of \( C^1 \) functions \( \{u^{(+)}\}_{i=1}^\infty \) and \( \{u^{(-)}\}_{i=1}^\infty \). By the continuity of the trace operator with respect to strict convergence in BV(\Omega) (see, e.g., [L. Theorem 3.88]), also \( u^i := u^{(+)} + i + u^{(-)} + i \), then converge strictly to \( u \). Moreover, \( D^s u^i = Du^i|\Gamma \). By (L.2) we therefore have
\[ |D\gamma#u^i|\Omega) + |D\gamma#u^i|\Omega) - 2|Du^i|\Omega) \leq G_{\gamma,\gamma}[D^s\gamma#u^i]U) + (|D\gamma#u^i|\Omega) + |D\gamma#u^i|\Omega) - 2|Du^i|\Omega) \]
Since the traces of \( u^i \) on \( \Gamma \) converge in \( L^1(\Gamma) \) to the trace of \( u \) on \( \Gamma \), we deduce the separable double-Lipschitz property by analogous arguments as the (non-separable) double-Lipschitz property above.

\[ \square \]

**Remark 4.1.** Observe that we obtained the estimate
\[ \text{TV}(\gamma#u) + \text{TV}(\gamma#u) - 2\text{TV}(u) \leq G_{\gamma,\gamma}[Du](U) \]
that involves none of \( D^\gamma \gamma^\perp \) and \( \text{TV}(\gamma#u) \). If we take \( \gamma(x) = \iota(x) := x \), then the property shows
\[ |D\gamma#u|\Omega) \leq (1 + G_{\gamma,\gamma})|Du|\Omega) \]
This can in specific cases can improve upon standard estimates [L. on the total variation under Lipschitz transformations.
4.2. Huber-regularised TV

For a parameter \( \eta > 0 \), Huber-regularised total variation may be defined as

\[
TV_\eta(u) := \sup \left\{ \int_\Omega u \div \varphi \, dx - \frac{\eta}{2} \| \varphi \|_{L^2(\Omega; \mathbb{R}^m)}^2 \mid \varphi \in C_c^\infty(\Omega; \mathbb{R}^m), \| \varphi \|_{2, L^\infty(\Omega; \mathbb{R}^m)} \leq 1 \right\}. \tag{4.4}
\]

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

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

giving

\[
TV_\eta(u) = \int_\Omega |\nabla u(x)|_\eta \, dx, \quad (C^1(\Omega)).
\]

Huber-regularisation of TV is sometimes helpful numerically, especially in the context of second-order optimisation methods \cite{Huber}, as well as primal-dual methods for non-convex problems \cite{Chambolle}. 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_\eta 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 limited co-coercive energies.

**Definition 4.1.** We denote by \( \Psi \) 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

\[
\psi^\infty := \lim_{t \to \infty} \psi(t)/t.
\]

Secondly, \( \psi \) is limited co-coercive in the sense that there exist constants \( K_{\psi}, C_{\psi} > 0 \) with

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

For a justification of the term “co-coercive”, we refer to \cite[Theorem 18.15]{Evans}. In essence, the “limited” co-coercivity contained in the threshold \( K_\psi \) says that TV_\psi, defined after the next lemma, behaves linearly like TV for large gradients. This of course holds for Huber-TV.

**Lemma 4.2.** We have \( | \cdot |_\eta \in \Psi, \ (\eta > 0) \).

**Proof.** Only the limited co-coercivity 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 co-coercivity in this range. In the case \( s < 1/\eta < t \), we also have

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

Limited co-coercivity follows with \( K_\psi = 1/\eta \) and \( C_\psi = \eta \).

**Definition 4.2.** Let \( \psi \in \Psi \). Then we set

\[
TV_\psi(u) := \int_\Omega \psi(\|\nabla u(x)\|) \, dx + \psi^\infty |D^* u|(\Omega), \quad (u \in BV(\Omega)).
\]

**Remark 4.2.** It can be shown that

\[
TV_\psi(u) = \sup \left\{ \int_\Omega u(x) \div \varphi(x) \, dx - \int_\Omega \psi^*(\|\varphi(x)\|) \, dx \mid \varphi \in C_c^\infty(\Omega; \mathbb{R}^m), \|\varphi\|_{2, L^\infty(\Omega; \mathbb{R}^m)} \leq 1 \right\}. \tag{4.5}
\]
We now claim that $\text{TV}_\psi$ satisfies the requirements of our main results, Theorem 4.2 and Theorem 4.3.

**Proposition 4.2.** Let $\Omega \subset \mathbb{R}^m$ be a bounded domain with Lipschitz boundary and $\psi \in \Psi$. Then $\text{TV}_\psi$ 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 (4.3).

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

**Lemma 4.3.** Let $\psi \in \Psi$ and $A, B : \mathbb{R}^m \to \mathbb{R}^m$ be linear transformations, and $c, d > 0$. Denote

$$
\tilde{G}_{A,B} := \sup_{\|v\|=1} \|Av\| + \|Bv\| - 2\|v\|,
$$

$$
\tilde{J}_{c,d} := |c + d - 2|, \text{ and }
$$

$$
\tilde{D}_A := \|A - I\|.
$$

If $\tilde{D}_A, \tilde{D}_B \leq M$ for some constant $M > 0$, then there exists a constant $C = C(M, K_\psi, \psi^\infty)$ such that

$$
c\psi(\|Av\|) + d\psi(\|Bv\|) - 2\psi(\|v\|) \leq C(\tilde{G}_{cA,dB} + \tilde{J}_{c,d} + \tilde{D}_A^2 + \tilde{D}_B^2\|v\|), \quad (v \in \mathbb{R}^m).
$$

**Proof.** Let us first of all observe directly from the definition of the subdifferential, and $\psi$ being increasing that

$$
0 < \sup_t \|\partial \psi(t)\| \leq \psi^\infty,
$$

(4.7)

From convexity also

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

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

$$
\psi(s) \leq \psi^\infty s, \quad (s \geq 0).
$$

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

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

(4.8)

If $\|Av\|, \|Bv\| \geq \|v\|$, we immediately deduce using (4.7) that

$$
L \leq \psi^\infty (c\|Av\| - d\|Bv\| - 2\|v\|) + (c + d - 2)(\psi(\|v\|) - \psi^\infty \|v\|)
$$

$$
\leq \psi^\infty (\tilde{G}_{cA,dB}\|v\| + \tilde{J}_{c,d}\|v\|).
$$

This is what we need. If $\|Av\|, \|Bv\| \leq \|v\|$, we deduce

$$
L \leq \psi^\infty \tilde{J}_{c,d}\|v\|.
$$

Again the claim holds. It remains to consider the case $\|Av\| \geq \|v\| > \|Bv\|$, the case with $A$ and $B$ exchanged being analogous. We now use the co-compactness of $\psi$ as follows. We pick $z_B \in \partial \psi(\|Bv\|)$ and $z_A \in \partial \psi(\|Av\|)$, and define

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

Then

$$
z_B(\|Bv\| - \|v\|) = (z_B - z_A)(\|Bv\| - \|v\|) + z_A(\|Bv\| - \|v\|)
$$

$$
= (z_B - z_A)(\|Bv\| - \|Av\|) + (z_B - z_A)(\|Av\| - \|v\|) + z_A(\|Bv\| - \|v\|)
$$

$$
\leq \tilde{\chi}C_\psi(\|Bv\| - \|Av\|)^2 + z_A(\|Bv\| - \|v\|).
$$

In the final step we have used the limited co-compactness, $\|Av\| - \|v\| \geq 0$, and the fact that $z_B - z_A \leq 0$ which follows from the monotonicity of $\partial \psi$. Thus, continuing from (4.8), we calculate

$$
L \leq c\tilde{\chi}(\|Av\| - \|v\|) + d\tilde{\chi}(\|Bv\| - \|v\|) + (c + d - 2)\psi(\|v\|)
$$

$$
\leq c\tilde{\chi}(\|Av\| - \|v\|) + dz_A(\|Bv\| - \|v\|) + (c + d - 2)\psi(\|v\|) + d\tilde{\chi}C_\psi(\|Bv\| - \|Av\|)^2
$$

$$
\leq \psi^\infty (\tilde{G}_{cA,dB} + \tilde{J}_{c,d})\|v\| + 2\tilde{\chi}C_\psi(\tilde{D}_A^2 + \tilde{D}_B^2)\|v\|^2.
$$
Finally, we observe that
\[ \|v\| \leq C \]
for some constant \( C = C(M, K_\psi) \). This proves the claim. \( \square \)

**Remark 4.3.** We only needed the limited co-coercivity and the bound \( M \) on \( \tilde{D}_A \) and \( \tilde{D}_B \) for the estimate \( \|v\|^2 \leq C\|v\| \) below \( K_\psi \).

**Proof of Proposition 4.3.** Weak* lower semicontinuity is immediate from the formulation (4.3). To see (4.4), we first of all observe for some constant \( c > 0 \) that
\[ \psi^*(z) \leq c(1 + \|z\|^2). \]
Indeed, for large \( t, s > 0 \), limited co-coercivity implies \( \partial \psi(s) = \partial \psi(t) = k \) for some constant \( k > 0 \). Therefore the definition of \( \psi^* \) gives the claimed bound.

Let us then pick \( \varphi \in C_c^\infty(\Omega; \mathbb{R}^m) \) with \( \|\varphi\|_{2,L^\infty(\Omega)} \leq 1 \). We estimate
\[ \int_\Omega u \div \varphi \, dx - \int_\Omega \psi^*(\|\varphi(x)\|) \, dx \geq \int_\Omega u \div \varphi \, dx - 2cL^m(\Omega). \]
It follows
\[ \text{TV}_\psi(\Omega) \geq |Du(\Omega)| - 2cL^m(\Omega). \]
This shows 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 \( |D^{\alpha} \gamma \# u(\Omega)| \) is unaltered in (4.1). In place of the absolutely continuous part \( |D^{\alpha} \gamma \# u(\Omega)| \), we have \( \int_\Omega \psi(\|\gamma \# u\|) \, dx \). Using the area formula, we calculate
\[ \int_\Omega \psi(\|\gamma \# u(x)\|) \, dx = \int_U \psi(\|\gamma^{-1}(\gamma(x))\|) J_{\gamma}(x) \, dx. \]
Let now \( \gamma, \gamma \in \mathcal{F}(\Omega, U) \) with \( T_{\gamma, \gamma} < 1 \). Summing the previous equation for \( \gamma = T_{\gamma, \gamma} \), subtracting \( 2 \int_\Omega \psi(\|\gamma \# u\|) \, dx \) and using Lemma 3.3 with \( A = \nabla^{-1}(\gamma(x)) \), \( B = \nabla^{-1}(\gamma(x)) \), \( c = J_{\gamma}(x) \), \( d = J_{\gamma}(x) \), and \( v = \nabla u(x) \) yields
\[ \int_\Omega \psi(\|\gamma \# u\|) \, dx + \int_\Omega \psi(\|\gamma \# u\|) \, dx - 2 \int_\Omega \psi(\|\gamma \# u\|) \, dx \leq C(G_{\gamma, \gamma} + J_{\gamma, \gamma} + D_{\gamma} + D_{\gamma}^2) \int_U \|\gamma \# u\| \, dx. \]
With this estimate at hand, the rest follows exactly as in Proposition 4.1.

Finally, to see, \( \|u\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)} \) for solutions \( u \) to (3), we follow the corresponding proof for TV. Namely, if we set \( \tilde{L} := \|f\|_{L^\infty(\Omega)} \), and replace \( u \) by
\[ \tilde{u}(x) := \max\{-L, \min\{u(x), L\}\}, \]
then it follows from the chain rule in BV [3, Theorem 3.96] that
\[ |D\tilde{u}(A)| \leq |Du(A)| \quad \text{for any Borel set } A. \]
From this it is immediate that
\[ \text{TV}_\psi(\tilde{u}) \leq \text{TV}_\psi(u), \]
Moreover, if \( \tilde{u} \neq u \), i.e., \( \|\tilde{u}\|_{L^\infty(\Omega)} < \|u\|_{L^\infty(\Omega)} \), we deduce
\[ \int_\Omega \phi(\tilde{u}(x) - f(x)) \, dx < \int_\Omega \phi(u(x) - f(x)) \, dx. \]
This provides a contradiction. Thus we must have \( \|u\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)} \). \( \square \)
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 \( \psi \). These however do not immediately form well-posed problems due to lack of weak* lower semicontinuity in \( BV(\Omega) \).

Remark 4.4 (Non-convex total variation). Regularisation functionals \( TV_{\psi} \) based on concave energies \( \psi \) have recently received increased attention, for the better modelling of real image gradient statistics. Let us define

\[
\psi^0 := \lim_{t \to 0} \psi(t)/t,
\]

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 \( \psi^0 \) 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 \( TV_{\psi} \) is still (separably) double-Lipschitz comparable. Indeed, Lemma 4.3 is trivial to prove in this case. Assuming for simplicity that \( c + d = 2 \), by concavity

\[
c\psi(||Av||) + d\psi(||Bv||) - 2\psi(||v||) \leq c\partial\psi(||v||)(||Av|| - ||v||) + d\partial\psi(||v||)(||Bv|| - ||v||)
\]

\[
\leq \psi^0 \max\{0, ||cAv|| + ||dBv|| - ||v||\} 
\]

\[
\leq \psi^0 \tilde{G}_{A,B} ||v||.
\]

The assumption \( c + d = 2 \) is also not difficult to remove with the help of \( J_{c,d} \), and is satisfied by the Lipschitz transformation we construct in Section 5.

Unfortunately, the \( TV_{\psi} \) regularisation functional constructed with concave \( \psi \) is not admissible. It lacks weak* lower semicontinuity. Using area strict convergence and additional multiscale regularisation, first introduced in [13] for discontinuous optical flow, problems involving such regularisation can however be made well-posed. It is outside the scope of the present paper to prove that the multiscale regularisation is separably double-Lipschitz comparable. Nevertheless, assuming we have a solution \( u \in BV(\Omega) \) to the basic model, which is not guaranteed, then our technique shows \( H^m_{f-1}(J_u \setminus J_f) = 0 \).

Remark 4.5 (Perona-Malik). The Perona-Malik anisotropic diffusion may also be written in the variational form with \( \phi(t) = t^2 \) and \( R(u) = TV_{\psi}(u) \) for \( \psi(t) = \log(1 + t^2) \). Observe that this \( \psi \) 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 [22]. Nevertheless, our approach can be adapted to study it, assuming we have a solution \( u \in BV(\Omega) \) to the problem. The existence is however generally not guaranteed. Indeed, with the notation of Lemma 4.3 by properties of the logarithm, we have

\[
c\psi(||Av||) = c \log(1 + ||Av||^2) = c \log(c + c||Av||^2) - c \log c.
\]

By the concavity of the logarithm, we have

\[
c \log(c + c||Av||^2) - c \log(1 + ||v||^2) \leq \frac{c}{1 + ||v||^2} (c||Av||^2 - ||v||^2 + c - 1)
\]

\[
= \frac{1}{1 + ||v||^2} ((c||Av||^2 - ||v||^2 + (c - 1)(||v||^2 + 1)).
\]

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

\[
c \log(c + c||Av||^2) + d \log(d + d||Bv||^2) - 2 \log(1 + ||v||^2) \leq \frac{1}{1 + ||v||^2} G_{cA,dB}'||v||^2,
\]

where

\[
G_{cA,dB}' = ||c^2 A^* A + d^2 B^* B - 2I||.
\]

With \( c + d = 2 \), we also have

\[
-c \log c - d \log d \leq 0.
\]

Therefore, using \( 1 + ||v||^2 \geq ||v|| \) and referring back to (4.3), we obtain

\[
c\psi(||Av||) + d\psi(||Bv||) - 2\psi(||v||) \leq G_{cA,dB}'||v||.
\]

This is not exactly the estimate of Lemma 4.3, but in Section 5, we will in fact estimate \( \tilde{G}_{cA,dB} \) through \( G_{cA,dB}' \); see Lemma 4.3. Therefore we may follow this reasoning with the argument of Proposition 4.2. We also assumed \( c + d = 2 \), but it would not be difficult to remove this assumption with the help of \( J_{c,d} \). The specific Lipschitz transformations that we construct next in Section 5 however do satisfy \( c + d = 2 \), that is, \( J_m \gamma(x) + J_m \gamma(x) = 2 \).
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 (\( \square \)) that does not satisfy \( H^{m-1}(J_u \setminus J_f) = 0 \). The idea is to take a Lipschitz graph \( \Gamma \) on \( J_u \) containing a violating point \( x_0 \), and then to move the jump forward or backward by an amount \( \rho \), 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 [40]. Before the construction, we provide a simple general lemma stating one of the most crucial parts of our technique.

5.1. A crucial estimate

We will use the following simple lemma to simplify the estimation of the double-Lipschitz comparison factor \( G_{\gamma_{\gamma}} \).

**Lemma 5.1.** We have

\[
G_{\gamma_{\gamma}} \leq \sup_{x \in \Omega} \frac{1}{2} \| [A_\gamma(x)]^* A_\gamma(x) + [A_\gamma(x)]^* A_\gamma(x) - 2I \|.
\]

**Proof.** By the concavity of the square root, for any \( 0 \leq v \in \mathbb{R}^m \), 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

\[
\| A_\gamma(x)v \| + \| A_\gamma(x)v \| - 2\|v\| \leq \frac{1}{2\|v\|} \langle v, ([A_\gamma(x)]^* A_\gamma(x) + [A_\gamma(x)]^* A_\gamma(x) - 2I)v \rangle
\]

\[
\leq \frac{1}{2} \| [A_\gamma(x)]^* A_\gamma(x) + [A_\gamma(x)]^* A_\gamma(x) - 2I \| \| v \|.
\]

The case \( v = 0 \) being trivial, this concludes the proof. \( \square \)

5.2. The construction

For brevity, we now define \( v_x := P^\perp_{\gamma_1} x \) and \( t_x := \langle z_\Gamma, x \rangle \) to be the components of \( x \) on \( V_\Gamma \) and along \( z_\Gamma \), for a Lipschitz graph \( \Gamma \).

**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 \( h \in W^{1,\infty}_0(z_\Gamma^{-1} \cap B(0,1)) \) satisfies \(-s/3 \leq h \leq s/3\). Define the transformation \( \gamma_h : U \to U \) by

\[
\gamma_h(x) := v_x + \tilde{\gamma}_{s,h}(t_x; v_x) z_\Gamma,
\]

where

\[
\tilde{\gamma}_{s,h}(t; v) := \begin{cases} 
  t + \frac{h(v_x) + t f_r(v)}{s}, & f_r(v) - s < t < f_r(v), \\
  t + \frac{h(v_x) - t f_r(v)}{s}, & f_r(v) \leq t < f_r(v) + s.
\end{cases}
\]

Then \( \gamma_h \) is \( 1 \)-to-1 and Lipschitz and

\[
J_m \gamma_h(x) = |\tilde{\gamma}_{s,h}(t_x; v_x)| = \begin{cases} 
  1 + \frac{h(v_x)}{s}, & f_r(v_x) - s < t_x < f_r(v_x), \\
  1 - \frac{h(v_x)}{s}, & f_r(v_x) \leq t_x < f_r(v_x) + s.
\end{cases}
\]

Then also exists a constant \( C = C(h, f_r) > 0 \) such that for \( \rho \in (-1,1) \), we have

\[
\begin{align*}
T_{\gamma_{s,h},\gamma_{s-h}} &\leq C\rho^2, & G_{\gamma_{s,h},\gamma_{s-h}} &\leq C\rho^2, & J_{\gamma_{s,h},\gamma_{s-h}} &= 0, & D_{\gamma_{s,h}} &\leq C|\rho|, \\
T_{\gamma_{s,h},\gamma_{s-h}}^{-1} &\leq C|\rho|, & G_{\gamma_{s,h},\gamma_{s-h}}^{-1} &\leq C|\rho|, & J_{\gamma_{s,h},\gamma_{s-h}}^{-1} &\leq C|\rho|, & D_{\gamma_{s,h}}^{-1} &\leq C|\rho|.
\end{align*}
\]

(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 (5.4) are in that case a bit more work to prove, and having such a transformation would only help very little with Lemma 7.3.

Proof. After rotation and translation, if necessary, we may assume w.log that $z_T = (0, \ldots, 0, 1)$, so that $x = (v, t) := (\eta, t)$. Then

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

and

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

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

To calculate the Lipschitz factor and the determinant, we study the eigenvalues $\lambda_1, \ldots, \lambda_m$ of

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

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

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

We have

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

Therefore, as claimed

$$
\mathcal{J} \gamma_h(x) = \sqrt{\prod_{i=1}^m \lambda_i} = \sqrt{\lambda_1(x) \lambda_2(x)} = |d_h| = |\gamma_h'(t; v)|.
$$

It remains to consider the bounds (5.3). Letting $\overline{\gamma} := \gamma_{ph}$ and $\gamma := \gamma_{-ph}$, and recalling that

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

by Lemma 4.1 we have

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

We want to estimate this further. As in the proof of Lemma 4.1, we have $\nabla \gamma^{-1}(\gamma_h(x)) = [\nabla \gamma_h(x)]^{-1}$ (a.e.). With

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

it can easily be verified that

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

It follows that

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

and

$$
[A_{\gamma_h}(x)]^* A_{\gamma_h}(x) = \begin{pmatrix} d_h^2 I + c_h^2 c_h & -c_h \\ -c_h^T c_h & 1 \end{pmatrix}.
$$
We observe that
\[
\begin{align*}
c \mp h &= \pm \rho c_h, \quad \text{and} \\
d_{\rho h}^2 &= 1 \pm 2 \rho d_h + \rho^2 d_h^2,
\end{align*}
\]
(5.4)  (5.5)
Thus \(d_{\rho h}^2 + d_{-\rho h}^2 - 2 = 2 \rho^2 d_h^2\), and we deduce
\[
[A_\gamma(x)]^* A_\gamma(x) + [A_{-\gamma}(x)]^* A_{-\gamma}(x) - 2I = 2 \rho^2 \left( d_{\rho h}^2 I + c_{\rho h}^T e_h \begin{pmatrix}
0 \\
0
\end{pmatrix} \right).
\]
It follows
\[
G_{\gamma, \gamma} \leq C \rho^2.
\]
Moreover, (5.3) also yields
\[
J_{\gamma_{\rho h}, \gamma_{-\rho h}} = d_{\rho h} + d_{-\rho h} - 2 = 0.
\]
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
\[
[\nabla \gamma_{\rho h}(x)]^{-1} - I = d_{\rho h}^{-1} \begin{pmatrix}
0 & 0 \\
-c_{\rho h}^T & 1 - d_{\rho h}
\end{pmatrix}
\]
and
\[
D_\gamma = \sup_{x \in \Omega} \| \nabla \gamma_{\rho h}^{-1}(\gamma_{\rho h}(x)) - I \| = \sup_{x \in \Omega} \| \nabla \gamma_{\rho h}(x) \|^{-1} - I \| \leq C \rho, \quad (0 < \rho < 1).
\]
This proves all the bounds in (5.2) involving both \(\gamma_{\rho h}\) and \(\gamma_{-\rho h}\). The bounds involving \(\iota\) 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
\[
J_m \gamma_h^{-1}(y) = 1/J_m \gamma_h(x) = 1/d_h,
\]
as well as \(\nabla \gamma_{h}(\gamma_h^{-1}(y)) = \nabla \gamma_h(x)\), so that
\[
A_{\gamma_h^{-1}} = \begin{pmatrix}
d_{\rho h}^{-1}I \\
-c_{\rho h}^T d_{\rho h}^{-1}
\end{pmatrix}.
\]
We skip the elementary details. \(\square\)

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 \(u \in BV(\Omega)\). Define
\[
M_\gamma := \sup_{x \in U} \| \gamma_h(x) - x \|.
\]
Then \(M_\gamma = \| h(v) \|_{L^\infty(\Ga)}\) and
\[
\int_U |u(\gamma_h(x)) - u(x)| dx \leq M_{\gamma_h} |Du|(U).
\]

**Proof.** Let us again, without loss of generality after rotation and translation, if necessary, that \(z_\Gamma = (0, \ldots, 0, 1)\), so that \(x = (v, t) := (v_x, t_x)\). Then
\[
\gamma_h(v, t) = (v, \gamma_v(t)),
\]
and the slice
\[
u(v)(t) := u(v, t), \quad (-s < t < s; v \in V_\Gamma)
\]
satisfies \(u_v \in BV(-r, r)\) for \(L^{m-1}\)-a.e. \(v \in V_\Gamma\). Thus
\[
|u_v(\gamma(t; v)) - u_v(t)| \leq |Du_v|(\{t, \gamma(t; v)\}) = \int_{-s}^s \chi_{\{t, \gamma(t; v)\}}(\tau) d|Du_v|(\tau).
\]
Here we use the convention \([a, b] := [b, a]\) if \(b < a\). Observe that \([t, \tilde{\gamma}(t; v)] \subset [t, t + M_{\gamma_h}]\) or \([t, \tilde{\gamma}(t; v)] \subset [t, t - M_{\gamma_h}]\). Using Fubini’s theorem and basic properties of one-dimensional slices in \(BV\), we may thus estimate
\[
\int_U |u(\gamma_h(x)) - u(x)| \, dx = \int_V \int_{-s}^s |u_v(\gamma_v(t)) - u_v(t)| \, dt \, d\mathcal{H}^{m-1}(v)
\]
\[
\leq \int_V \int_{-s}^s \int_{-s}^s X_{[t, \tilde{\gamma}(t; v)]}(\tau) \, d|Du_v| (\tau) \, dt \, d\mathcal{H}^{m-1}(v)
\]
\[
\leq \int_V \int_{-s}^s \int_{-s}^s X_{[t, \tilde{\gamma}(t; v)]}(\tau) \, d|Du_v| (\tau) \, dt \, d\mathcal{H}^{m-1}(v)
\]
\[
= \int_V \int_{-s}^s M_{\gamma_h} \, d|Du_v| (\tau) \, d\mathcal{H}^{m-1}(v)
\]
\[
= M_{\gamma_h} |Du|(U).
\]

Finally
\[
\|\gamma_h(x) - x\| = |h(v)| \begin{cases} \frac{s - f_t(v) + t}{s + f_t(v) - t}, & f_t(v) - s < t < f_t(v), \\ f_t(v) \leq t < f_t(v) + s. \end{cases}
\]

Clearly this gives \(M_{\gamma_h} = \|h(v)\|_{L^\infty(V_r)}\).

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 \(0 \neq \tilde{h} \in W_0^{1,\infty}(z_1^0 \cap B(0, 1))\) satisfying \(0 \leq \tilde{h} \leq 1\). Set
\[
h_r(v) := r \tilde{h}(v - v(x_0))/r, \quad (r > 0).
\]
Then there exists \(r_0 > 0\) such that each of the maps
\[
\gamma_{\rho, r}(x) := v_x + \tilde{\gamma}_{3r, h_r}(t_x; v_x) z_\Gamma, \quad (-1 < \rho < 1, 0 < r < r_0),
\]
satisfies \(\gamma_{\rho, r} \in \mathcal{F}(\Omega; U_r)\) for
\[
U_r := x_0 + z_1^0 \cap B(0, r) + (3 + \text{lip } f_t)(-r, r) z_\Gamma.
\]
Moreover, there exists a constant \(C > 0\) such that the transformations \(\gamma_{\rho, r}, (0 < r < r_0, -1 < \rho < 1)\), satisfy
\[
T_{\gamma_{\rho, r}} \leq C \rho^2, \quad T_{\gamma_{\rho, r}} \leq C |\rho|,
\]
\[
M_{\gamma_{\rho, r}} = |\rho| r, \quad T_{\gamma_{\rho, r}} \leq C |\rho|.
\]

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

**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 \(0 \leq \tilde{h} \leq 1\), evidently also \(M_{\gamma} \leq |\rho|r\), and \(\gamma\) reduces to the identity outside \(U_r\). The remaining claims follow by applying Lemma 5.4 on
\[
\Gamma' := \gamma_{\Omega}(V(r) \cap B(v(x_0), r)) \subset \Gamma
\]
with \(h = h_r\) and \(s = 3r\).

6. **Proof of the main result for \(\rho > 1\)**

We now begin the proof of Theorem 5.2. 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 Section 5.3 after a technical lemma in Section 5.4. The fidelity estimates are based on the specific Lipschitz transformations constructed in the previous section. Averaging over two different Lipschitz transformations \(\gamma = \gamma_{\rho, r}\) and \(\gamma = \gamma_{-\rho, r}\) will provide an \(O(\rho)\) decrease estimate for the fidelity. Importantly, in order to take advantage of the strict convexity of \(\phi\), we actually need as our improvement candidates convex combinations \(\theta u + (1 - \theta)\gamma u\). After dealing with the fidelity function, we apply in Section 5.2 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 5.2 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 $u \in \mathcal{BV}(\Omega)$, we now fix a countably family of Lipschitz $(m-1)$-graphs $\{\Gamma_i\}_{i=1}^\infty$, such that
\[ \mathcal{H}^{m-1}(J_u \setminus \bigcup_{i=1}^\infty \Gamma_i) = 0. \] (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 $x_0 \in \Gamma^{x_0} \in \Gamma_i$ for some $i = i(x_0)$, satisfying the following two properties. Firstly $V_{\Gamma^{x_0}} \subset B(P_{\Gamma_i} x_0, r(x_0))$ for some $r(x_0) > 0$. This can clearly be satisfied since we can by Kirzbraun’s theorem assume $V_\Gamma = z_1^\perp$. Secondly we can take the traces of $u$ from both sides of $\Gamma$ 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 $u \in \mathcal{BV}(\Omega)$. 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
\[ \Theta^*_m(|Du|_\nu, (\Gamma^x)^+; x) = 0, \quad \text{and} \quad \Theta^*_m(|Du|_\nu, (\Gamma^x)^-; x) = 0. \]

**Proof.** We define $A_i$ as the set of points $x \in \Gamma_i$ such that
\[ \Theta^*_m(|Du|_\nu, \Gamma_i^+; x) = 0, \quad \text{and} \quad \Theta^*_m(|Du|_\nu, \Gamma_i^-; x) = 0. \] (6.2)

As a consequence of the BV extension theorem [3, Proposition 3.21], there exist extensions $v_i^\pm$ of $u|\Omega \cap \Gamma_i^\pm$ over $\Gamma_i$, satisfying $\mathcal{H}^{m-1}(\Gamma_i \cap J_{v_i}) = 0$. From this fact, [3, and Proposition 3.21] we deduce
\[ \mathcal{H}^{m-1}(\Gamma_i \setminus A_i) = \mathcal{H}^{m-1}(\Gamma_i \cap A_i^\perp) \leq \mathcal{H}^{m-1}(\Gamma_i \cap J_{v_i}) = 0. \]

Next, we define $B_i$ as the set of points $x \in \Gamma_i$ such that $x$ is a Lebesgue point of the traces $u^+|\Gamma_i$ and $u^-|\Gamma_i$. Again $\mathcal{H}^{m-1}(\Gamma_i \setminus B_i) = 0$. If we now set
\[ Z_u := J_u \setminus \bigcup_{i=1}^\infty (A_i \cap B_i), \]
we deduce that $\mathcal{H}^{m-1}(Z_u) = 0$. \hfill $\square$

6.2. The effect of the shift transformation on the fidelity

We now fix $x_0 \in J_u \setminus (J_f \cup Z_u)$, and let $\Gamma = \Gamma^{x_0}$, recalling from the construction above that the traces of $u$ from both sides of $\Gamma$ 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 $\gamma \in W_0^{1,\infty}(z_1^\perp \cap B(0,1))$ with $0 \leq \gamma \leq 1$. (Sometimes we only assume $-1 \leq \gamma \leq 1$. This is explicitly stated.) We set $h_r(v) := rh((v - v_{x_0})/r)$, and define the Lipschitz transformations $\gamma_{\rho,r} : \Omega \to \Omega$ by
\[ \gamma_{\rho,r}(x) := v_x + \gamma_{3r, h_r}(tx_x; v_x) z_{1r}, \quad (-1 < \rho < 1, 0 < r < r_0). \]

These transformations are the identity outside
\[ U_r := x_0 + z_1^\perp \cap B(0,r) + (3 + \text{lip } f_r)(-r,r)2r, \]
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
\[ U_r \subset B(x_0, Cr) \quad \text{for} \quad C = \sqrt{1 + (3 + \text{lip } f_r)^2}. \]

We pick arbitrary $\theta \in [0,1]$ as well as $r \in (0,r_0)$ and $\rho \in (0,1)$. With these we define $\tilde{u}_{\rho,r}(x) = \theta u(x) + (1 - \theta)\gamma_{\rho,r} u(x)$,
as well as the piecewise constant functions
\[
\begin{align*}
  f_0 &= \tilde{f}(x_0), \\
  u_0 &= u^+(x_0)\chi_{U_+^+} + u^-(x_0)\chi_{U_-^-}, \quad \text{and} \\
  \bar{u}_{\rho,r,0} &= \theta u_0(x) + (1-\theta)\gamma_{\rho,r,\#}u_0(x).
\end{align*}
\]

We aim to prove that for suitable \((\theta, \rho, r)\), 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 \(\theta\) will in particular be necessary to take advantage of the \(p\)-increase (or strict convexity) of \(\phi\) 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 \(O(\varepsilon r^m)\) 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 \(O(\varepsilon r^m)\).

**Lemma 6.2.** Suppose \(\phi\) is \(p\)-increasing with \(1 \leq p < \infty\). Suppose either \(u, f \in BV(\Omega) \cap L^\infty_{\text{loc}}(\Omega)\), or that \(p = 1\) and \(u, f \in BV(\Omega)\). Let \(x_0 \in J_u \setminus (S_f \cup Z_u)\). Given \(\varepsilon > 0\), there exists \(r_0 > 0\), independent of \(\rho\) and \(\gamma\), valid for every \(h \in W^{1,\infty}_0(\mathbb{R}^d \cap B(0,1))\) with \(-1 \leq \tilde{h} \leq 1\), such that whenever \(0 < r < r_0\) and \(-1 < \rho < 1\), then
\[
\begin{align}
  \int_{\Omega} \phi(\tilde{u}_{\rho,r}(x) - f(x)) &\leq \int_{\Omega} \phi(u(x) - f(x)) + \varepsilon |\rho|^m.
  \end{align}
\]

**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 \(h\). 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 \(O(\varepsilon r^m)\) sets \(W_r^+ := U_r^+ \setminus \gamma(U_r^-)\) and \(W_r^- := U_r^- \setminus \gamma(U_r^+\)) and within the \(O(r^m)\) sets \(W_r^+_r := U_r^+ \cap \gamma(U_r^+)\) and \(W_r^-_r := U_r^- \cap \gamma(U_r^-)\). We begin with the latter, observing that \(\tilde{u}_{\rho,r,\#}(x) = u_0(x)\) for \(x \in W_r^+_r\). Thus
\[
\begin{align}
  \int_{W_r^+_r} \phi(\tilde{u}_{\rho,r,\#}(x) - f(x)) &= \phi(u_0(x) - f(x)) dx = 0.
\end{align}
\]

We estimate using Definition \[\text{3.3}\] for \(L^m\)-a.e. \(x \in W_r^+\) that
\[
\phi(\tilde{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x)) \leq \frac{C_\phi}{\varepsilon} |\tilde{u}_{\rho,r}(x) - f(x)| - |u(x) - f(x)| |u_{\rho,r}(x) - f(x)|^{p-1}
\]
\[
\leq C_\phi (1-\theta)|\gamma_{\rho,r,\#}u(x) - u(x)| |\tilde{u}_{\rho,r}(x) - f(x)|^{p-1}
\]
\[
\leq C |\gamma_{\rho,r,\#}u(x) - u(x)|.
\]

In the final inequality we use the boundedness of \(u\), or alternatively \(p = 1\). We now employ Lemma \[\text{3.2}\] to estimate
\[
\begin{align}
  \int_{W_r^+} \phi(\tilde{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x)) dx &\leq C \int_{W_r^+} |\gamma_{\rho,r,\#}u(x) - u(x)| dx
  \\
  &\leq CM_{\gamma_{\rho,r,\#}}^{-1} |Du|(U_r^+).
\end{align}
\]

We have \(M_{\gamma_{\rho,r,\#}}^{-1} \leq \rho r\). By the construction of \(Z_u\), because \(x_0 \in J_u \setminus Z_u\), choosing \(r > 0\) small enough, we can enforce \(|Du|(U_r^+) \leq \varepsilon r^{-m}/(8C)\). Thus
\[
\begin{align}
  \int_{W_r^+} \phi(\tilde{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x)) dx &\leq \varepsilon \rho r^m/8.
\end{align}
\]

It remains to estimate \(\phi(\tilde{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x))\) on \(W_r^-\). By Definition \[\text{3.2}\] we have
\[
-\phi(u(x) - f(x)) \leq -\phi(u_0(x) - f_0(x)) + C_\phi |\tilde{u}_{\rho,r,\#}(x) - u(x) - f(x)| |u_0(x) - f_0(x)|^{p-1}.
\]

Since \(x \mapsto |u_0(x) - f_0(x)|^{p-1}\) is bounded, it follows for some constant \(C > 0\) that
\[
\begin{align}
  -\phi(u(x) - f(x)) &\leq -\phi(u_0(x) - f_0(x)) + C |u_0(x) - f_0(x) - u(x) + f(x)|.
\end{align}
\]

We may estimate the final term in \(W_r^-\) by
\[
\begin{align}
  \int_{W_r^-} |u_0(x) - f_0(x) - u(x) + f(x)| dx &\leq \int_{W_r^-} |u^-(x_0) - u(x)| dx + \int_{W_r^-} |\tilde{f}(x_0) - f(x)| dx.
\end{align}
\]
We now recall the notation \( u^*_z(t) = u(x + tz) \) for one-dimensional restrictions of functions of bounded variation in a direction \( z \) with \( \|z\| = 1 \), as well as the formula

\[
|\langle Du, z \rangle(A)| = \int_{\mathbb{R}^+ \setminus A} |Du^*_z|((t \in \mathbb{R} \mid x + tz \in A)) \, dx.
\]

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

\[
\int_{W_r^\pm} |u^\pm(x_0) - u(x)| \, dx = \int_{\Gamma \cap U_r} \int_0^{\rho h_r^{\pm}(v_x)} |u^\pm(x_0) - u(x \pm tz_\Gamma)| \, dt \, dx
\]

\[
\leq \int_{\Gamma \cap U_r} \int_0^{\rho h_r^{\pm}(v_x)} |u^\pm(x) - u(x \pm tz_\Gamma)| + |u^\pm(x) - u^\pm(x_0)| \, dt \, dx
\]

\[
\leq \int_{\Gamma \cap U_r} \int_0^{\rho h_r^{\pm}(v_x)} |Du^\pm_\Gamma((x, tz_\Gamma))| \, dt \, dx + \rho \int_{\Gamma \cap U_r} |u^\pm(x) - u^\pm(x_0)| \, dx
\]

\[
\leq \rho |\int U_r^\pm| + \rho \int_{\Gamma \cap U_r} |u^\pm(x) - u^\pm(x_0)| \, dx.
\]

Since \( x_0 \in J_u \setminus Z_u \) is a Lebesgue point of \( u^\pm \) on \( \Gamma \) and \( \Theta_{m-1}^\ast(|Du||\Gamma^\pm; x_0) = 0 \), choosing \( r > 0 \) small enough, we can make the final quantity less than \( \rho r \cdot \epsilon r^{m-1}/(16C) \). Performing similar estimation of \( \tilde{f}(x_0) - f \), and referring back to (6.7), we deduce

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

It remains to estimate the transformed terms on \( W_r^\pm \). Similarly to (6.4), using the essential boundedness of both \( u \) and \( f \) on \( U_r \), or \( p = 1 \), we deduce for \( \mathcal{L}^m \)-a.e. \( x \in W_r^\pm \)

\[
\phi(\tilde{u}_{\rho,r}(x) - f(x)) \leq \phi(\tilde{u}_{\rho,r,0}(x) - f_0(x)) + C|\tilde{u}_{\rho,r,0}(x) - f_0(x) - \tilde{u}_{\rho,r}(x) + f(x)|. \tag{6.9}
\]

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

\[
\int_{W_r^\pm} \phi(\tilde{u}_{\rho,r}(x) - f(x)) \, dx \leq \int_{W_r^\pm} \phi(\tilde{u}_{\rho,r,0}(x) - f_0(x)) \, dx + \epsilon \rho r^m/8. \tag{6.10}
\]

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

**Lemma 6.3.** Suppose \( \phi \) is \( p \)-increasing and \( u, f \in L^\infty(U_r) \). Let \( \rho > 0 \) and

\[
K^\pm := \text{ess sup}_{x \in U_r}(|\tilde{u}_{\pm,r}(x) - f(x) - |u(x) - f(x)|).
\]

Then

\[
\int_{U_r} \phi(\tilde{u}_{\pm,r}(x) - f(x)) \, dx - \int_{U_r} \phi(u(x) - f(x)) \, dx \leq K^\pm C\phi \int_{U_r} |\tilde{u}_{\pm,r}(x) - f(x)|^{p-1} \, dx.
\]

**Proof.** By Definition 6.3, we have

\[
\phi(\tilde{u}_{\rho,r}(x) - f(x)) - \phi(u(x) - f(x)) \leq C\phi(|\tilde{u}_{\rho,r}(x) - f(x)| - |u(x) - f(x)|)|\tilde{u}_{\rho,r}(x) - f(x)|^{p-1}
\]

\[
\leq K^\pm C\phi|\tilde{u}_{\rho,r}(x) - f(x)|^{p-1}.
\]

The case with negative shift \(-\rho\) is proved analogously.

**Lemma 6.4.** Suppose \( \phi \) is \( p \)-increasing with \( p > 1 \). Then there exist \( \theta \in (0,1) \) and a constant \( C = C(\phi, u^\pm(x_0), \tilde{f}(x_0)) > 0 \) such that for \( \rho > 0 \) holds

\[
\int_{U_r} \phi(\tilde{u}_{\rho,r,0}(x) - f_0(x)) \, dx + \int_{U_r} \phi(\tilde{u}_{-\rho,r,0}(x) - f_0(x)) \, dx - 2 \int_{U_r} \phi(u_0(x) - f_0(x)) \, dx \leq -C\rho r^m.
\]
Proof. Let \( \rho > 0 \). Since \( \tilde{h} \geq 0 \), observe that outside \( \widehat{W}^\pm_r := U^\pm_r \cap \gamma_{\pm,r}(U^\mp_r) \) the transformed functions agree with the originals,

\[
\phi(\bar{u}_{\pm,r,0}(x) - f_0(x)) = \phi(u_0(x) - f_0(x)).
\]

Using Lemma 3.3 on \( \widehat{W}^\pm_r \) with \( u = u_0 \) and \( f = f_0 \) thus gives

\[
\int_{U_r} \phi(\bar{u}_{\pm,r,0}(x) - f_0(x)) \, dx - \int_{U_r} \phi(u_0(x) - f_0(x)) \, dx \leq K^\pm C\phi \int_{\widehat{W}^\pm_r} |\bar{u}_{\pm,r,0}(x) - f_0(x)|^{p-1} \, dx.
\]

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

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

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

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 6.1 in the appendix

\[
\mathcal{L}^m(\widehat{W}^+_r) = \mathcal{L}^m(\widehat{W}^-_r) = \int_{V_r} \rho |b_r(v)| \, dH^{m-1}(v) = C_h \rho r^m
\]

for some constant \( C = C_h \). Thus

\[
\int_{U_r} \phi(\bar{u}_{\pm,r,0}(x) - f_0(x)) \, dx + \int_{U_r} \phi(\bar{u}_{-\pm,r,0}(x) - f_0(x)) \, dx - 2 \int_{U_r} \phi(u_0(x) - f_0(x)) \, dx \leq C\phi C_h \rho r^m (K^+ \zeta^+ + K^- \zeta^-).
\]

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

\[
K^+ \zeta^+ + K^- \zeta^- < 0. \tag{6.11}
\]

We calculate

\[
K^+ \leq (1 - \theta)(|b^- - a| - |b^+ - a|), \tag{6.12}
\]

and

\[
K^- \leq (1 - \theta)(|b^+ - a| - |b^- - a|). \tag{6.13}
\]

Therefore

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

We concentrate on the case \( b^+ < b^- \), with \( a \geq (b^+ + b^-)/2 \), otherwise handled analogously by appropriate changes of roles and negations. Then

\[
|b^+ - a| \geq |b^- - a|. \tag{6.14}
\]

If (6.14) holds strictly, we have \( \zeta^+ > \zeta^- \) for \( \theta \in (0, 1) \) large enough. This shows (6.11) and is the only place where we need the assumption \( p > 1 \). If (6.14) 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 \( \theta \). 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 (6.11), and may conclude the proof of the lemma. \( \square \)

Lemma 6.5. Suppose \( \phi \) is \( p \)-increasing with \( 1 < p < \infty \), and both \( u, f \in BV(\Omega) \cap L^\infty_{\text{loc}}(\Omega) \). Let \( x_0 \in J_a \setminus (S_f \cup Z_a) \). Then there exist \( \theta \in (0, 1), r_0 > 0, \) independent of \( \rho \), and a constant \( C = C(\phi, u^\pm(x_0), \tilde{f}(x_0)) > 0 \), such that whenever \( 0 < r < r_0 \) and \( 0 < \rho < 1 \), it holds

\[
\int_{\Omega} \phi(\bar{u}_{\pm,r}(x) - f(x)) \, dx + \int_{\Omega} \phi(\bar{u}_{-\pm,r}(x) - f(x)) \, dx - 2 \int_{\Omega} \phi(u(x) - f(x)) \, dx \leq -C \rho r^m. \tag{6.15}
\]

Proof. Combine Lemma 3.3 and Lemma 6.4, choosing \( \epsilon > 0 \) small enough in the former. \( \square \)
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 Section 6.2.

Lemma 6.6. 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

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

Proof. We know from Lemma 5.2 the existence of a constant $C > 0$ such that

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

By convexity

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

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

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

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

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

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 5.2. Suppose, to reach a contradiction, that $\mathcal{H}^{m-1}(J_u \setminus J_f) > 0$. Since also $\mathcal{H}^{m-1}(Z_u) = 0$, we may select a point $x_0 \in J_u \setminus (J_f \cup Z_u)$. We then apply Lemma 6.3 for the fidelity estimate

$$\int_{\Omega} \phi(\bar{u}_{\rho,r}(x) - f(x)) \, dx + \int_{\Omega} \phi(\bar{u}_{-\rho,r}(x) - f(x)) \, dx - 2\int_{\Omega} \phi(u(x) - f(x)) \, dx \leq -C_1 \rho r^m, \quad (6.16)$$

where $0 < r < r_1$ and $0 < \rho < 1$. Lemma 6.6 gives for the regulariser and $0 < r < r_2$ the estimate

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

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

$$\left(\int_{\Omega} \phi(\bar{u}_{\rho,r}(x) - f(x)) \, dx + R(\bar{u}_{\rho,r})\right) + \left(\int_{\Omega} \phi(\bar{u}_{-\rho,r}(x) - f(x)) \, dx + R(\bar{u}_{-\rho,r})\right) - 2\left(\int_{\Omega} \phi(u(x) - f(x)) \, dx + R(u)\right)$$

$$\leq C_2 \rho^2 r^{-m-1} - C_1 \rho r^m.$$  

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

\[ \boxed{ } \]
7. Considerations for the $L^1$ fidelity

The only difficulty in extending the proofs of Section 6 for $p$-increasing ($p > 1$) fidelities $\phi$ to the $L^1$ fidelity $\phi(x) = x$, is Lemma 6.4. We do not necessarily get a constant $C > 0$ there, but $C = 0$. Then the argument in the proof of Theorem 6.3 does not go through. This has the consequence that it is possible to have $H^{m-1}(J_u \setminus J_f) > 0$. 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 $TV$. To begin with the study, we provide alternative versions of some of the technical lemmas of Section 6.

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

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

for some $0 \neq \tilde{h} \in W^{1,\infty}_0(z_f^+ \cap B(0,1))$, with $0 \leq \tilde{h} \leq 1$ unless explicitly stated otherwise. However, we also denote

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

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

$$ I_r := \int_{V_r} h_r \, dv, $$

and

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

7.1. Estimate of the fidelity

We begin with the modifications to Lemma 6.3.

Lemma 7.1. Suppose $\phi(t)$ is 1-increasing, and that $u, f \in BV(\Omega)$. 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 $\tilde{h} \in W^{1,\infty}_0(z_f^+ \cap B(0,1))$ with $-1 \leq \tilde{h} \leq 1$, we have

$$ \int_{\Omega} \phi(\bar{u}_{\rho,r}(x) - f(x)) \, dx - \int_{\Omega} \phi(u(x) - f(x)) \, dx 
\leq \epsilon \rho r^m - \sigma(1 - \theta)\rho C_\phi \lambda_u(x_0) I_r, \quad (0 < r < r_0, -1 < \rho < 1). \tag{7.1} $$

Proof. We assume that $\rho \geq 0$; the case $\rho < 0$ can be handled by negating $\tilde{h}$. 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 (7.1) for the latter piecewise constant functions.

Referring to Definition 6.3 we deduce

$$ \phi(\bar{u}_{\rho,r,0}(x) - f_0(x)) - \phi(u_0(x) - f_0(x)) \leq C_\phi (|\bar{u}_{\rho,r,0}(x) - f_0(x)| - |u_0(x) - f_0(x)|) |\bar{u}_{\rho,r,0}(x) - f_0(x)|^0 $$

This is non-zero only in $W^+_r := U_r^+ \cap \gamma_{\rho,r,0}(U_r^-)$, and in $W^-_r := U_r^- \cap \gamma_{\rho,r}^0(U_r^+)$. For $x \in W^+_r$, as in the proof of Lemma 6.3,

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

Here we again denote $b^+ := u^+(x_0)$ and $a := f(x_0)$. 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

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

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 \( h_r^+, h_r^- \geq 0 \), thus
\[
L := \int_{U_r} \phi(\bar{u}, u, 0(x) - f_0(x)) - \phi(u_0(x) - f_0(x)) \, dx
\]
\[
\leq C\phi K^+ \zeta^+ \int_{V_r} \rho^h_r - C\phi K^- \zeta^- \int_{V_r} \rho^h_r(v) \, dv.
\]
As in the proof of Lemma 5.4, we concentrate on the case \( b^+ < b^- \) with \( a \geq (b^+ + b^-)/2 \), setting \( \sigma = 1 \). Other cases are handled analogously with \( \sigma = -1 \). Since \( b^+ \neq b^- \), there are at most two choices of \( \theta \) for which \( \zeta^+ = 0 \) or \( \zeta^- = 0 \). Simply by the triangle inequality
\[
K^- \leq (1 - \theta)|b^+ - b^-|.
\]
Further, choosing \( \theta < 1 \) large enough while maintaining \( \zeta^+ = 1 \), we can ascertain
\[
K^+ \leq (1 - \theta)(b^+ - b^-) = -(1 - \theta)|b^+ - b^-|.
\]
Thus
\[
L \leq C\phi|b^+ - b^-| \left( \int_{V_r} \rho^h_r(v) \, dv - \int_{V_r} \rho^h_r(v) \, dv \right) = -C\phi \lambda u(x_0) I_r. \]

**Remark 7.1.** Looking at the proof of Lemma 5.4 with a little bit more care in the approximation of \( K^- \) above, we could in fact get the stronger double-sided estimate (5.17) if \( u \) “jumps through \( f \), in other words, if
\[
\min\{u^+(x), u^-(x)\} < \tilde{f}(x_0) < \max\{u^+(x), u^-(x)\}.
\]

### 7.2. Estimates for the regulariser

We now start work on the counterpart of Lemma 5.4. The proof will use a different method, and only one-sided Lipschitz estimates. We begin with the missing technical curvature definition required for Theorem 3.3.

**Definition 7.1.** Let \( R \) be an admissible regularisation functional on \( BV(\Omega) \). We define the transformation differential of \( R \) at \( u \) as
\[
\mathcal{D}^R_u(\gamma) := \lim_{\rho \to 0} \frac{R((\rho \gamma + (1 - \rho) u) \# u) - R(u)}{\rho}, \quad (\gamma \in \mathcal{F}(\Omega)),
\]
when the limit exists. With \( \bar{h} \in W^{1, \infty}_0(z^0 I \cap B(0, 1)) \), we then define the pointwise \( R \)-curvature at \( x_0 \) along the Lipschitz graph \( \Gamma \), if the limit exists, as
\[
\mathcal{C}^R_{u, \Gamma, \bar{h}}(x_0) := \lim_{r \to 0} I_r^{-1} \mathcal{D}^R_u(\gamma_{h_r}), \quad (x_0 \in \Gamma).
\]
Since our results do not depend on the choice of \( \bar{h} \), we simply write \( \mathcal{C}^R_{u, \Gamma}(x_0) := \mathcal{C}^R_{u, \Gamma, \bar{h}}(x_0) \) in the statement of Theorem 3.3.

**Remark 7.2.** In many ways, it would make more sense to define the pointwise curvature by
\[
\tilde{\mathcal{C}}^R_{u, \Gamma, \bar{h}}(x_0) := \lim_{r \to 0} \omega_{m-1} \mathcal{D}^R_u(\gamma_{1, r}) / I_r |\mathcal{D}u|([B(x_0, r])).
\]
At a point \( x_0 \in J_u \) this would give the normalised curvature \( \tilde{\mathcal{C}}^R_{u, \Gamma, \bar{h}}(x_0) / \lambda u(x_0) \); see Proposition 5.1. 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.

**Proposition 7.1.** 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) \( \mathcal{C}^R_{u, \Gamma, \bar{h}}(x_0) = \alpha^{-1} \).

(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
\[
R(\gamma_{\rho_{j, i}, r_j} \# u) - R(u) - \kappa \alpha^{-1} \rho_{j, i} I_r \leq 0, \quad (7.2)
\]
or
\[
R(\gamma_{\rho_{j, i}, r_j} \# u) - R(u) - \kappa^{-1} \alpha^{-1} \rho_{j, i} I_r \geq 0. \quad (7.3)
\]
Proof. We observe that $\gamma_{\rho,x} = \rho \gamma_{1,r} + (1 - \rho)\iota$, so that
\[
D_u R(\gamma_{1,r}) = \limsup_{\rho \to 0} \frac{R(\rho \gamma_{1,r} + u) - R(u)}{\rho}.
\]
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 $\rho' > 0$ such that for every $0 < \rho < \rho'$ we have bounds
\[
\kappa^{-1} \alpha^{-1} \rho I_r \leq R(\rho \gamma_{1,r} + u) - R(u) \leq \kappa \alpha^{-1} \rho I_r.
\]  
(7.4)
If we define the upper and lower transformation differentials
\[
D_u R^+(\gamma) := \limsup_{\rho \to 0} \frac{R(\gamma \# u) - R(u)}{\rho} \quad \text{and} \quad D_u R^-(\gamma) := \liminf_{\rho \to 0} \frac{R(\gamma \# u) - R(u)}{\rho},
\]
then dividing (7.4) by $\rho$ and letting $\rho \to 0$, it follows that
\[
\kappa^{-1} \alpha^{-1} I_r \leq D_u R^+(\gamma_{1,r}) \leq D_u R^-(\gamma_{1,r}) \leq \kappa \alpha^{-1} I_r.
\]
Dividing by $I_r$ and letting $r \to 0$, we find
\[
\kappa^{-1} \alpha^{-1} \leq \liminf_{r \to 0} I_r^{-1} D_u R^+(\gamma_{1,r}) \leq \limsup_{r \to 0} I_r^{-1} D_u R^-(\gamma_{1,r}) \leq \kappa \alpha^{-1}
\]
Since $\kappa \in (0, 1)$ was arbitrary, we deduce
\[
\alpha^{-1} = C_u R, \Gamma, h(x_0).
\]
In particular, $C_u R, \Gamma, h(x_0)$ exists and case (II) holds. \hfill \qed

7.3. Regularity

We now study the regularity of $J_f \setminus J_u$ at points where Proposition [I.1] holds. We assume that $Du$ is "essentially piecewise constant" around such points, as mainly employed in the density assumption of the next lemma.

Lemma 7.2. Let $\Gamma \subset \Omega$ be a Lipschitz $(m - 1)$-graph and $x_0 \in \Gamma$. Suppose $R$ is separably double-Lipschitz comparable, and $\Theta_m(|Du|, \Omega \setminus \Gamma; x_0) = 0$. Then
\[
C_u R, \Gamma, h(x_0) = \lim_{r \to 0} I_r^{-1} \overline{D}_u R(\gamma_{1,r}),
\]  
(7.5)
where
\[
\overline{D}_u R(\gamma_{1,r}) := \lim_{\rho \to 0} R^s \frac{|D \gamma_{1,r \# u}((\gamma_{1,r}(\Gamma)) - |Du|^{(\Gamma)})}{\rho}.
\]
Proof. We recall that
\[
D_u R(\gamma_{1,r}) = \lim_{\rho \to 0} \frac{R(\gamma_{1,r \# u}) - R(u)}{\rho}.
\]
By Definition [I.3] we have
\[
R(\gamma_{1,r \# u}) - R(u) \leq R^s T_{\gamma_{1,r \# u}}(|Du|)^2 U \setminus \Gamma + R^s (|D |Du|^{(\Gamma)}).
\]
Using Lemma [I.2] to estimate $T_{\gamma_{1,r \# u}}$, and minding the assumption $\Theta_m(|Du|, \Omega \setminus \Gamma; x_0) = 0$, we deduce for arbitrary $\epsilon > 0$, for small enough $r > 0$ that
\[
D_u R(\gamma_{1,r}) \leq \epsilon |Du| + R^s \limsup_{\rho \to 0} \frac{|D \gamma_{1,r \# u}((\gamma_{1,r}(\Gamma)) - |Du|^{(\Gamma)})}{\rho}.
\]
Dividing by $I_r$, letting $r \to 0$, and afterwards letting $\epsilon \to 0$, yields the upper bound
\[
C_u R, \Gamma, h(x_0) \leq \lim_{r \to 0} I_r^{-1} \overline{D}_u R(\gamma_{1,r}).
\]
In order to derive the corresponding lower bound, we set $\gamma := \gamma_{1,r}$ and $v := \gamma \# u$. Then
\[
R(\gamma \# u) - R(u) = R(v) - R(\gamma^{-1} \# v) \geq -R^s T_{\gamma^{-1}, u}(|Du|)^2 U \setminus \gamma(\Gamma) - R^s |D |Du|^{(\Gamma)}.
\]
Using Lemma [I.4] to estimate $T_{\gamma^{-1}, u}$, and proceeding as above, we deduce (7.5). \hfill \qed
Lemma 7.3. Let $\Gamma \subset \Omega$ be a Lipschitz $(m-1)$-graph. At $H^{m-1}$-a.e. point $x_0 \in \Gamma$, if $\Theta_m(\{Du\}, \Omega \setminus \Gamma; x_0) = 0$, then
\[
C^R_{u, \Gamma, h}(x_0) = -\frac{\nabla f_{\Gamma}(P_{x_0}h)}{1 + \|\nabla f_{\Gamma}(P_{x_0}h)\|^2} \cdot R^s \lambda_u(x_0).
\]

Proof. We define the shorthand notation $\lambda(v) := \lambda_u(\gamma_\Gamma(v))$, and observe from Lemma 7.2 in the appendix that
\[
|D\gamma_{\rho, r}u|_2(\gamma_{\rho, r}(\Gamma)) = \int_{V_\rho} \lambda(v) \sqrt{1 + \|\nabla f_{\Gamma} + \rho h_r(v)\|^2} dv,
\]
and
\[
|Du|_2(\Gamma) = \int_{V_\rho} \lambda(v) \sqrt{1 + \|\nabla f_{\Gamma}\|^2} dv.
\]
We further calculate using convexity of the norm, and concavity of the square root that
\[
\rho c_{\lambda, f}(h_r) \leq \int_V \lambda(v) \sqrt{1 + \|\nabla f_{\Gamma} + \rho h_r(v)\|^2} dv - \int_V \lambda(v) \sqrt{1 + \|\nabla f_{\Gamma}\|^2} dv
\]
\[
\leq \rho c_{\lambda, f}(h_r) + \rho^2 \int_V \lambda(v) \frac{\|\nabla h_r(v)\|^2}{\sqrt{1 + \|\nabla f_{\Gamma}\|^2}} dv,
\]
where
\[
c_{\lambda, f}(h) := \int_V \lambda(v) \left( \frac{\nabla f_{\Gamma}(v)}{\sqrt{1 + \|\nabla f_{\Gamma}(v)\|^2}}, \nabla h(v) \right) dv.
\]
Thus
\[
\tilde{D}^R_{u, \Gamma, h}(\gamma_{1, r}) = R^s c_{\lambda, f}(h_r).
\]
Minding Lemma 7.2, we deduce
\[
C^R_{u, \Gamma, h}(x_0) = \lim_{r \to 0} R^s c_{\lambda, f}(h_r/I_r).
\]
Since $\nabla f_{\Gamma}$ is bounded, the claim is immediate if $\gamma_{\rho, r}(x_0)$ is a Lebesgue point of $\lambda$ and $\nabla f_{\Gamma}$, that is $H^{m-1}$-a.e. \hfill $\Box$

Lemma 7.4. Let $f \in W^{1, \infty}(V)$, and $0 \leq \lambda \in L^\infty(V)$, where $V \subset \mathbb{R}^m \cap B(0, r_0)$, $z \in \mathbb{R}^m$, $r_0 > 0$. Suppose for some $\alpha > 0$ we have
\[
-\text{div} \frac{\nabla f(v)}{\sqrt{1 + \|\nabla f(v)\|^2}} = \alpha^{-1}, \quad (7.6)
\]
for $H^{m-1}$-a.e. $v \in V$. Then (7.6) holds for every $v \in V$, and $f \in C^{2, \alpha}(V)$.

Proof. Since (7.6) holds almost everywhere, also
\[
\int_V \left( \frac{\nabla f(v)}{\sqrt{1 + \|\nabla f(v)\|^2}}, \nabla h(v) \right) dv = \alpha^{-1} \int_V h(v) dv, \quad (h \in W^{1, \infty}_0(V)). \quad (7.7)
\]
After showing $C^2$ regularity of $f$, (7.4) 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
\[
\min \{ \mathcal{J}(q) \mid q \in Q \}, \quad (7.8)
\]
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 $h \in W^{1, \infty}_0(V)$ to the differential equation
\[
\nabla h = \nabla \tilde{h} \text{ on } V, \quad h_{\partial V} = 0,
\]
is unique for any $\tilde{h} \in W^{1, \infty}_0(V)$, we find that $\mathcal{J}$ is strictly convex on $Q$. It therefore has a unique minimiser. But (7.7) 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 $\tilde{\mathcal{J}}$. 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

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

where

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

and

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

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

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

$$\rho \mapsto \hat{J}(q + \rho h), \quad (h \in W^{1,\infty}_0(V)), $$

we calculate

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

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

$$\int_V (\nabla g(\nabla q(x)), \nabla h(x)) \, dx = \alpha^{-1} \int_V h(x) \, dx, \quad (h \in W^{1,\infty}_0(V)),$$

where

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

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

$$\text{div} A(\nabla q) = -\alpha^{-1}, \quad q \in W^{1,\infty}(V), \quad q|\partial V = f|\partial V, \quad (7.9)$$

where

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

We want to show that $\hat{J}$ has a solution $\hat{q} \in C^{2,\lambda}(V)$ for any $\lambda \in (0, 1)$. But $\hat{J}$ had the unique minimiser $f$. Therefore $f = \hat{q} \in C^{2,\lambda}(V)$ for any $\lambda \in (0, 1)$. From (7.8) and the definition of $\psi$ we have moreover that (7.6) holds.

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

$$A_i \in C^{1,\lambda}(\mathbb{R}^m), \quad (i = 1, \ldots, m), \quad (7.10)$$

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

$$\langle p, A(p) \rangle \geq \|p\|^{2+\tau} - C_1, \quad (7.11)$$

and

$$\|\nabla A(p)\| \leq C_2 (1 + \|p\|^\tau). \quad (7.12)$$

It is to force these conditions, why we introduced the $\psi$-penalty in $g$. Minding that $\psi \in \mathcal{C}^3(\mathbb{R})$, we have $A \in \mathcal{C}^{2}(\mathbb{R}^m, \mathbb{R}^m)$, whence condition (7.10) 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 (7.11) holds with \( \tau = 2 \). It remains to show that (7.14) 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 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 (7.12), concluding the proof.

7.4. 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. The general result is contained in the next lemma, and afterwards in the proof of Theorem 3 itself, we study additional regularity under approximate piecewise constancy.

**Lemma 7.5.** Suppose \( u \) solves (1). Let \( \Gamma \subset \Omega \) be a Lipschitz \((m-1)\) graph. Then \( C_u^{R,\Gamma,\tilde{h}}(x_0) = \lambda_u(x_0) C_\phi \) for \( H^{m-1} \) a.e. \( x_0 \in (J_u \setminus J_f) \cap \Gamma \).

**Proof.** By Lemma 4.1, given \( \epsilon > 0 \), we have the existence of \( r_0 > 0 \) and \( \sigma \in \{-1, +1\} \) such that

\[
\int_{\Omega} \phi(u_{\rho, r}(x) - f(x)) \, dx - \int_{\Omega} \phi(u(x) - f(x)) \, dx \leq \epsilon \rho^m - \sigma (1 - \rho) \rho C_\phi \lambda_u(x_0) I_{r}, \quad (0 < r < r_0, -1 < \rho < 1). \tag{7.13}
\]

We concentrate on the \( \sigma = +1 \) case; the case \( \sigma = -1 \) is handled analogously (or simply by exchanging sides \( \Gamma^+ \) and \( \Gamma^- \)). We suppose first that Proposition 7.1[(i)] 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 (7.14) holds. Then for some \( \kappa \in (0, 1) \), a sequence \( r_{i} \searrow 0 \), and for each fixed \( j \), a sequence \( \rho_{j,i} \searrow 0 \), we have

\[
R(\bar{u}_{\rho_{j,i}, r_{i}}) - R(u) \leq (1 - \theta) \left( R(\gamma_{\rho_{j,i}, r_{i}, \bar{u}}) U - R(u) \right) \leq (1 - \theta) \kappa C_\phi \rho_{j,i} \lambda_u(x_0) I_{r_{i}}.
\]

Taking \( r = r_{j} \) and \( \rho = \rho_{j,i} \) and summing with (7.13), we find that

\[
\left( \int_{\Omega} \phi(u_{\rho_{j,i}, r_{j}}(x) - f(x)) \, dx + R(\bar{u}_{\rho_{j,i}, r_{j}}) \right) - \left( \int_{\Omega} \phi(u(x) - f(x)) \, dx + R(u) \right) \leq \epsilon \rho_{j,i} r_{j}^m - (1 - \kappa)(1 - \theta) \rho_{j,i} C_\phi \lambda_u(x_0) I_{r_{j}} \leq \epsilon \rho_{j,i} r_{j}^m - C'' \rho_{j,i} r_{j}^m.
\]

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 (1).

Suppose then that (7.3) holds. Then for some \( \kappa > 0 \) we have

\[
R(\bar{u}_{\rho_{j,i}, r_{j}}) - R(u) \geq (1 - \theta) \kappa^{-1} C_\phi \rho_{j,i} \lambda_u(x_0) I_{r_{j}}. \tag{7.14}
\]

We now use the double-Lipschitz comparability we already used in the proof of Lemma 4.6. Namely, choosing \( r > 0 \) small enough, we have for some constant \( C' > 0 \) that

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

We may w.l.o.g assume that all \( \{r_{i}\}_{i \in \mathbb{Z}^+} \) are small enough for this. Using (7.14) it follows

\[
R(\bar{u}_{\rho_{j,i}, r_{j}}) - R(u) \leq C' \rho_{j,i} r_{j}^{-m-1} - (1 - \theta) \kappa^{-1} C_\phi \rho_{j,i} \lambda_u(x_0) I_{r_{j}}. \tag{7.15}
\]

Summing (7.13) with (7.14) we find that

\[
\left( \int_{\Omega} \phi(u_{\rho_{j,i}, r_{j}}(x) - f(x)) \, dx + R(\bar{u}_{\rho_{j,i}, r_{j}}) \right) - \left( \int_{\Omega} \phi(u(x) - f(x)) \, dx + R(u) \right) \leq C' \rho_{j,i} r_{j}^{-m-1} + \epsilon \rho_{j,i} r_{j}^m + (1 - \kappa^{-1})(1 - \theta) \rho_{j,i} C_\phi \lambda_u(x_0) I_{r_{j}} \leq C' \rho_{j,i} r_{j}^{-m-1} + \epsilon \rho_{j,i} r_{j}^m - C'' \rho_{j,i} r_{j}^m.
\]

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 (1). It follows that Proposition 7.1[(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 \( C_u^{R,\Gamma,\tilde{h}}(x_0) = \lambda_u(x_0) C_\phi \) for \( H^{m-1} \) a.e. \( x_0 \in (J_u \setminus J_f) \cap \Gamma \).
Proof of Theorem 7.4. Let \( \{ \Gamma_i \}_{i=1}^{\infty} \) be the collection of Lipschitz graphs from (7.3). Repeated application of Lemma 7.3 for each \( \Gamma = \Gamma_i, (i \in \mathbb{Z}^+) \) shows that

\[
C_{u}^{R, \Gamma, h}(x_0) = \lambda_u(x_0)C_\phi \quad \text{for} \quad \mathcal{H}^{m-1}\text{-a.e. } x_0 \in (J_u \setminus J_f) \cap \Gamma_i.
\]

Choosing \( \Lambda_i = \Gamma_i \) proves (7.8).

It remains to prove the higher regularity (7.4) under assumptions of approximate piecewise constancy. We plan to use Lemma 7.3 and Lemma 7.4. The use of Lemma 7.4 on \( \Gamma = \Lambda_i \) depends on \( C^{R, \Lambda_i, h}(x_0) = \lambda_u(x_0)C_\phi \) holding \( \mathcal{H}^{m-1}\text{-a.e. on } \Lambda_i \), and the domain \( V_{\Lambda_i} \) being open. We therefore need a tiny extra argument before its application.

We fix \( \Gamma = \Gamma_i \) for some \( i \in \mathbb{Z}^+ \). By assumption \( \Theta_m(\|Du\|_\Omega \setminus \Gamma; x_0) = 0 \) at \( \mathcal{H}^{m-1}\text{-a.e. } x_0 \in \Gamma \). By Lemma 7.3 we therefore have at \( \mathcal{H}^{m-1}\text{-a.e. } x_0 \in \Gamma \) the expression

\[
C_{u}^{R, \Gamma, h}(x_0) = -\text{div} \frac{\nabla f_{\Gamma}(P_{z_i}^+x_0)}{\sqrt{1 + \|\nabla f_{\Gamma}(P_{z_i}^+x_0)\|^2}} \cdot \lambda_u(x_0).
\]

Alternatively

\[
C_{u}^{R, \Gamma, h}(x_0) = \lim_{r \searrow 0} c_{1,f}(h_r/I_r) \cdot \lambda_u(x_0).
\]

where we recall from the proof of Lemma 7.3 that

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

The functional \( c_{1,f} \) is continuous on \( H^1_0(V_{\Gamma}) \). 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 \( V_{\Gamma} \). 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 \( \text{ri} W_k^\sigma \neq \emptyset \). However \( \mathcal{H}^{m-1}(\text{ri} W_k^\sigma \cap (J_u \setminus J_f)) = 0 \), because otherwise we would find a set of positive mass of points \( x_0 \in J_u \setminus J_f \) where Proposition 7.3 does not hold, and reach a contradiction. Because we could always replace \( \Gamma \) by \( \Gamma \setminus g_{\Gamma}((\text{ri} W_k^\sigma) \setminus v) \), we may therefore assume that \( \text{ri} W_k^\sigma = \emptyset \) for every \( k \in \mathbb{Z}^+ \) and \( \sigma > 0 \). In particular every \( W_k^\sigma \) is nowhere dense in \( V_{\Gamma} \). The Baire category theorem then says that

\[
V_{\Gamma}^\sigma := \bigcap_{k=1}^{\infty} \bigcap_{j=1}^{\infty} (V_{\Gamma} \setminus W_k^{1/j}) = \bigcap_{k=1}^{\infty} \bigcap_{j=1}^{\infty} \bigcup_{0 < r < 1/j} (V_{\Gamma} \setminus W_{k,r}).
\]

is a dense open set within \( V_{\Gamma} \). But every \( v \in V_{\Gamma}^\sigma \) satisfies for some \( r_j \searrow 0 \) the limit

\[
\lim_{j \to \infty} c_{1,f}(h_r^v/I_{r_j}) = C_\phi.
\]

Considering that \( \lim_{r_j \searrow 0} c_{1,f}(h_r^v/I_{r_j}) \) exists at a Lebesgue point of \( \nabla f_{\Gamma} \), we deduce that \( C_{u}^{R, \Gamma, h}(g_{\Gamma}(v)) = \lambda_u(g_{\Gamma}(v))C_\phi \) for \( \mathcal{H}^{m-1}\text{-a.e. } v \in V_{\Gamma}^\sigma \). Moreover, as required

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

for \( \Lambda = \bigcup_{i=1}^{\infty} \Lambda_i \) with \( \Lambda_i := g_{\Gamma}(V_{\Lambda_i}) \). 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 \( \Lambda_i \) under the assumption that \( \Theta_m(\|Du\|_\Omega \setminus \Lambda_i; x) = 0 \) at \( \mathcal{H}^{m-1}\text{-a.e. point } x \in \Lambda_i. \)

\[\square\]
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.

References

1. W. Allard, Total variation regularization for image denoising, I. Geometric theory, SIAM Journal on Mathematical Analysis 39 (2008), 1150–1190.
2. F. Alter, V. Caselles and A. Chambolle, A characterization of convex calibrable sets in $\mathbb{R}^n$, Mathematische Annalen 332 (2005), 329–366.
3. L. Ambrosio, N. Fusco and D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Oxford University Press, 2000.
4. H. Bauschke and P. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics, Springer, 2011.
5. M. Benning and M. Burger, Ground states and singular vectors of convex variational regularization methods, Methods and Applications of Analysis 20 (2013), 295–334. [arXiv:1211.2057]
6. 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), 764–790, doi:10.1002/cpa.20011
7. K. Bredies, K. Kunisch and T. Pock, Total generalized variation, SIAM Journal on Imaging Sciences 3 (2011), 492–526, doi:10.1137/090769521
8. K. Bredies, K. Kunisch and T. Valkisch, Properties of $L^1$-TGV$^2$: The one-dimensional case, Journal of Mathematical Analysis and Applications 398 (2013), 438–454. doi:10.1016/j.jmaa.2012.08.053, URL http://math.uni-graz.at/mobis/publications/SFB-Report-2011-006.pdf
9. M. Burger, M. Franek and C.-B. Schönlieb, Regularized regression and density estimation based on optimal transport, Applied Mathematics Research eXpress 2012 (2012), 209–253, doi:10.1093/amrx/abs007
10. 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), 879–894.
11. A. Chambolle, An algorithm for mean curvature motion, Interfaces and Free Boundaries 6 (2004), 195.
12. A. Chambolle and P.-L. Lions, Image recovery via total variation minimization and related problems, Numerische Mathematik 76 (1997), 167–188.
13. T. Chan, A. Marquina and P. Mulet, High-order total variation-based image restoration, SIAM Journal on Scientific Computation 22 (2000), 503–516, doi:10.1137/S1064827598334169
14. T. F. Chan and S. Esedoglu, Aspects of total variation regularized $L^1$ function approximation, SIAM Journal on Applied Mathematics (2002), 1817–1837.
15. T. F. Chan, S. H. Kang and J. Shen, Euler’s elastica and curvature-based inpainting, SIAM Journal on Applied Mathematics (2004), 564–592.
16. 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), 1–26.
17. 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), 2351–2391, doi:10.1137/070697822
18. S. Didas, J. Weickert and B. Burgeth, Properties of higher order nonlinear diffusion filtering, Journal of Mathematical Imaging and Vision 35 (2009), 208–226, doi:10.1007/s10851-009-0166-3
19. V. Duval, J. F. Aujol and Y. Gousseau, The TVL1 model: A geometric point of view, Multiscale Modeling and Simulation 8 (2009), 154–189.
20. I. Fonseca and G. Leoni, Modern methods in the calculus of variations: $L^p$ spaces, Springer Verlag, 2007.
21. D. Gilbarg and N. Trudinger, Elliptic partial differential equations of second order, Classics in mathematics, Springer, 2001.
22. P. Guidotti, Anisotropic diffusions of image processing from Perona-Malik on, Advanced Studies in Pure Mathematics (2014). To appear.
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 $0 \leq h \in W_{0}^{1,\infty}(V_{\Gamma})$, 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

\[ \mathcal{L}_m(\Gamma^- \cap \Gamma^+) + \mathcal{L}_m(\Gamma_\rho^{+} \cap \Gamma^-) = |\rho| \int_{V_\Gamma} h(v) \, dv, \quad (A.1) \]

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

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

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

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

Analogously for \( \rho < 0 \) we obtain

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

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 (2.1) 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 (2.1) 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

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