



**HAL**  
open science

# Inclusion and estimates for the jumps of minimizers in variational denoising

Antonin Chambolle, Michal Łasica

► **To cite this version:**

Antonin Chambolle, Michal Łasica. Inclusion and estimates for the jumps of minimizers in variational denoising. *SIAM Journal on Imaging Sciences*, 2023, 17 (3), pp.1844-1878. 10.1137/23M1627948 . hal-04323807v1

**HAL Id: hal-04323807**

**<https://hal.science/hal-04323807v1>**

Submitted on 5 Dec 2023 (v1), last revised 25 Nov 2024 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Inclusion and estimates for the jumps of minimizers in variational denoising

Antonin Chambolle<sup>1</sup> and Michał Lasica<sup>2</sup>

<sup>1</sup>CEREMADE, CNRS, Université Paris-Dauphine, PSL University, Paris,  
MOKAPLAN, INRIA, Paris, France

<sup>2</sup>Institute of Mathematics of the Polish Academy of Sciences, Warsaw, Poland

chambolle@ceremade.dauphine.fr    mlasica@impan.pl

December 5, 2023

## Abstract

We study in this note the stability and inclusion of the jump set of minimizers of convex denoising functionals, such as the celebrated “Rudin–Osher–Fatemi” functional, for scalar or vectorial signals. We show that under mild regularity assumptions on the data fidelity term and the regularizer, the jump set of the minimizer is essentially a subset of the original jump set. Moreover, we give an estimate on the magnitude of jumps in terms of the data. This extends old results, in particular of the first author (with V. Caselles and M. Novaga) and of T. Valkonen, to much more general cases. We also consider the case where the original datum has unbounded variation, and define a notion of its jump set which, again, must contain the jump set of the solution.

**Keywords** Inverse problems, variational methods, total variation, imaging.

## 1 Introduction

The Total Variation regularizer was proposed for image denoising in [36] and has become popular for its simplicity and its ability to recover edges and discontinuities in the restored images. Even if it is largely outdated and has much lower performances than non-local [11, 7, 28], (learned) patches and dictionary-based [16, 42] or neural network based [21] techniques, it remains useful as a regularizer for large scale inverse problems (sometimes combined with machine learning and plug-n-play type [40] methods, see for instance [41]), as it is convex and relatively simple to optimize, in particular in combination with other (ideally also convex) terms.

An interesting question, answered first in [8], is whether a total variation-denoising method can create spurious structures and discontinuities, or if the edge set of the original image is preserved. Precisely, given  $f \in BV(\Omega)$  a (scalar) function with bounded variation, representing the grey-level values of an image defined in a domain  $\Omega \subset \mathbb{R}^m$  ( $m$  an integer, 2 or 3 in most applications), and with jump  $J_f$ , one considers  $u$  which solves:

$$\min_u \int_{\Omega} |Du| + \frac{1}{2} \int_{\Omega} |u - f|^2 dx. \quad (1)$$

The main result of [8] asserts that  $J_u \subset J_f$  (up to a negligible set for the surface measure); in addition,  $u^+ - u^- \leq f^+ - f^-$  a.e. on the jump set of  $u$ . It is also deduced that the  $L^2$ -gradient flow of the total variation, starting from an initial function  $u(0) \in L^{m/(m-1)}(\Omega)$ , has a diminishing jump set:  $s > t > 0 \Rightarrow J_{u(s)} \subset J_{u(t)}$  (more precise results are found in [10]). This is generalized to some integrands (such as the graph area, anisotropic total variations) already in [8], and further variants (including strictly convex data terms) in [23, 24]; see also [22]. The approach in the above mentioned papers consists in comparing the curvatures of the level sets of minimizers. One shows that at (approximate) continuity points of  $f$ , these curvatures are determined by the level, and ordered in a way which excludes the possibility that the boundaries of two different level sets coincide. The technique is relatively simple and elegant, and even allows to derive basic regularity results away from the jumps [31, 9], but it is restricted to the scalar case.

An alternative approach was proposed by T. Valkonen in [38]. It relies only on regularity properties of the regularizer, and therefore is not limited to the scalar case. In particular, the case where  $u, f$  are vector valued and the total variation is defined by means of the Frobenius norm of the matrix  $Du$  should enter the framework of [38], even if this does not seem to be explicit in the literature. One reason for this is the relative complexity of the criterion in [38] (the *double-Lipschitz comparability condition*), which is not always straightforward to check in practice, and the technicality of the proofs which most probably limited the audience of the papers [38, 39], despite their interest and originality.

In this new study, we introduce a general approach for addressing the issue of jump inclusion and control in total variation denoising and similar variational problems. Essentially, we show that jump inclusion occurs when the regularizer is differentiable with respect to an elementary class of inner variations of the solution. We also derive an estimate on the size of the jump (see for instance [10]). The approach applies to many regularizers, such as the Frobenius or (more surprisingly) the Nuclear (or Trace) Norm-based total variation in a vectorial setting<sup>1</sup> (see for instance [19, 15]). More interestingly, while the extension of Valkonen's approach to higher order regularizers, addressed in [39], excludes the "Total Generalized Variation" (TGV) of [6], a relatively simple modification of our proof allows to show jump inclusion in a slightly regularized version of that case, at least whenever the solution  $u$  is bounded (which can be enforced by a box type constraint in the minimization). The result for the exact "TGV" case remains open and, if true, probably requires a mix of our techniques and the ideas in [39], which address successfully other types of inf-convolution based regularizers.

Our approach is based on a very simple observation: at a jump point, the data term (such as the squared norm in (1)) will have different left and right derivatives along inner variations orthogonal to the jump, so that, if the regularizer is differentiable, some inequality is derived which involves only the data term. The idea can be illustrated by an elementary 1D example: consider  $\Omega = ]-1, 1[ \subset \mathbb{R}$ ,  $f \in BV(\Omega)$ , and let  $u$  minimize (1). Consider then  $\bar{x} \in J_u$ , with  $u^+(\bar{x}) > u^-(\bar{x})$ . Without loss of generality we assume  $u^+(\bar{x})$  is the right-sided limit of  $u$  at  $\bar{x}$ . Denote  $f^+(\bar{x})$  the right-sided limit of  $f$ , and  $f^-(\bar{x})$  the left-sided limit (with possibly  $f^+(\bar{x}) \leq f^-(\bar{x})$ ). Then, if  $\varphi$  is a smooth approximation of  $\chi_{[\bar{x}-\delta, \bar{x}+\delta]}$ , for  $\delta > 0$  small, for

---

<sup>1</sup>Precisely, our results hold for Schatten-type norms, but not Ky-Fan type norms such as the Spectral Norm.

$\tau \in ]0, \delta[$  one has:

$$\int_{\Omega} (u(x + \tau\varphi(x)) - f(x))^2 - (u(x) - f(x))^2 dx \approx \tau[(u^+(\bar{x}) - f^-(\bar{x}))^2 - (u^-(\bar{x}) - f^-(\bar{x}))^2],$$

$$\int_{\Omega} (u(x - \tau\varphi(x)) - f(x))^2 - (u(x) - f(x))^2 dx \approx \tau[(u^-(\bar{x}) - f^+(\bar{x}))^2 - (u^+(\bar{x}) - f^+(\bar{x}))^2]$$

as  $\tau \rightarrow 0$ . On the other hand, since the total variation of  $u(x \pm \tau\varphi(x))$  is the same as the total variation of  $u$ , thanks to minimality of  $u$  in (1) we deduce, sending  $\tau \rightarrow 0$ :

$$(u^+(\bar{x}) - f^-(\bar{x}))^2 - (u^-(\bar{x}) - f^-(\bar{x}))^2 \geq 0,$$

$$(u^-(\bar{x}) - f^+(\bar{x}))^2 - (u^+(\bar{x}) - f^+(\bar{x}))^2 \geq 0,$$

that is:  $(u^+(\bar{x}) - u^-(\bar{x}))(u^+(\bar{x}) + u^-(\bar{x}) - 2f^-(\bar{x})) \geq 0$  and  $(u^+(\bar{x}) - u^-(\bar{x}))(u^+(\bar{x}) + u^-(\bar{x}) - 2f^+(\bar{x})) \leq 0$ . We deduce that

$$f^-(\bar{x}) \leq \frac{u^+(\bar{x}) + u^-(\bar{x})}{2} \leq f^+(\bar{x})$$

(and in particular  $f^-(\bar{x}) \leq f^+(\bar{x})$ ), so that either  $\bar{x} \in J_f$ , or  $(u^+ + u^-)/2 = f$  at  $\bar{x}$ . This is elementary, and almost the conclusion we would like to reach.

Actually proving the jump set inclusion (and an estimate on the jump) in any dimension, following the same idea, is not much harder but requires a more subtle choice of the variation. The solution is found in Valkonen's work [38, Sec. 6], which uses a competitor for the minimization problem given by a convex combination of the minimizer itself and its inner variation, see Lemma 2 below. We show here (by a much simpler argument/calculation than in [38]) that together with the differentiability of the regularizer, it is enough to get a general estimate on the jump of  $u$ . This is done in Section 3 (Theorem 1).

Further (Sec. 4), we discuss general regularizers which satisfy the assumptions for our main result to hold. In particular, we find that the Frobenius or Nuclear-norm based Total Variations for vectorial-valued images meet our differentiability criterion (Section 4.2). In Section 5 we discuss conditions which ensure that the solution  $u$  to our variational problems, in the unconstrained case, are locally bounded. In Section 6 we show how a small adjustment of the proof extends the result to the inf-convolution type regularizers such as smoothed variants of the Total Generalized Variation (TGV) [6], studied in [39] again with a more complicated approach, and only partial conclusions.

Section 7 is devoted to the case where the data term is not necessarily of bounded variation: in that case, we introduce a notion of jump (as the set of points where the function differs significantly on both sides of a hyperplane) for which we still can show that it must contain the jump set of the solution. Interestingly, this should apply to data term consisting of the sum of a  $BV$  function and a bounded oscillating noise, such as many examples of noisy images. We provide, as an illustration, such an example in Section 8 and show the reconstruction with various types of color total variations.

## 2 Preliminaries

### 2.1 General notation

We will consider  $\mathbb{R}^n$ -valued functions,  $n \geq 1$ , defined on some open subset of  $\mathbb{R}^m$ ,  $m \geq 1$  (most of the proofs are written for  $m \geq 2$ , yet the case  $m = 1$  follows by trivial simplification). Given

$x \in \mathbb{R}^m$ ,  $r > 0$ ,  $\nu \in \mathbb{S}^{m-1}$  a unit vector, one denotes:

$$\begin{aligned} B_r(x) &= \{y \in \mathbb{R}^m : |y - x| < r\}, & B_r^\pm(x, \nu) &= \{y \in B_r(x) : \pm \nu \cdot (y - x) \geq 0\}, \\ B_r^{m-1}(x, \nu) &= B_r(x) \cap (x + \nu^\perp), & Q_r(x, \nu) &= B_r^{m-1}(x, \nu) + ]-r, r[\nu, \\ Q_r^+(x, \nu) &= B_r^{m-1}(x, \nu) + [0, r[\nu, & Q_r^-(x, \nu) &= B_r^{m-1}(x, \nu) + ]-r, 0]\nu. \end{aligned} \quad (2)$$

where  $|\cdot|$  is the standard Euclidean norm.

For  $\mu$  a Radon measure and  $k \leq m$ , we define for any  $x \in \Omega$  the  $k$ -dimensional density of  $\mu$  at  $x$  as the limit:

$$\Theta^k(\mu, x) = \lim_{r \rightarrow 0} \frac{\mu(B_r(x))}{\omega_k r^k},$$

when it exists. Here,  $\omega_k$  is the volume of the unit ball of dimension  $k$ .

## 2.2 The approximate discontinuity set and the jump set

Let  $w \in L^1_{loc}(\Omega)^n$ . Following [3, Definition 3.63], we say that  $w$  has an approximate limit at  $x \in \Omega$  if there exists  $z \in \mathbb{R}^n$  such that

$$\lim_{r \rightarrow 0^+} \int_{B_r(x)} |w(y) - z| dy = 0. \quad (3)$$

If no such  $z$  exists,  $x$  is called an approximate discontinuity point of  $w$ . The set of all approximate discontinuity points of  $w$  is denoted  $S_w$ . It is well known that  $\mathcal{L}^m(S_w) = 0$  [3, Proposition 3.64].

On the other hand, if there exist  $\nu_w \in \mathbb{S}^{m-1}$ ,  $w^\pm \in \mathbb{R}^n$ ,  $w^- \neq w^+$ , such that

$$\lim_{r \rightarrow 0^+} \int_{B_r^\pm(x, \nu_w)} |w(y) - w^\pm| dy = 0, \quad (4)$$

$x$  is called an (approximate) jump point of  $w$ . The set of all jump points of  $w$  is called the (approximate) jump set of  $w$  and is denoted by  $J_w$ . Clearly  $J_w \subset S_w$ . However, the condition defining jump points is rather rigid: even for a general locally integrable function, the jump set is countably  $\mathcal{H}^{m-1}$ -rectifiable [13]—that is, it can be covered up to a  $\mathcal{H}^{m-1}$ -negligible set by a countable union of Lipschitz or, equivalently,  $C^1$  graphs [3, p. 80].

We observe that  $B_r^\pm(x, \nu_w)$  may be replaced with  $Q_r^\pm(x, \nu_w)$  in (4) without changing the definition. Moreover, if  $x \in \Omega \setminus S_w$ , then (4) holds with  $w^+ = w^- = z$  and any  $\nu_w \in \mathbb{S}^{m-1}$ . Thus, (4) defines a (multi)function  $x \mapsto \{w^+(x), w^-(x)\}$  on  $J_w \cup (\Omega \setminus S_w)$ . For  $x \in J_w$ , the triple  $(w^+(x), w^-(x), \nu_w(x))$  is defined uniquely up to a permutation of  $(w^+, w^-)$  and a change of sign of  $\nu_w$ . In particular, the tensor product  $(w^+(x) - w^-(x)) \otimes \nu_w(x)$  is uniquely defined for  $x \in J_w$  (and for  $x \in \Omega \setminus S_w$ , where it vanishes).

We recall the notion of Lebesgue points closely related to approximate continuity. If  $\mu$  is a Radon measure on  $\Omega$  and  $w \in L^p_{loc}(\Omega, \mu)^n$ ,  $p \in [1, \infty[$ , we say that  $x \in \Omega$  is a ( $p$ -)Lebesgue point of  $w$  (with respect to  $\mu$ ), if

$$\lim_{r \rightarrow 0^+} \int_{B_r(x)} |w(y) - w(x)|^p d\mu(y) = 0.$$

It is known that  $\mu$ -almost every  $x \in \Omega$  is a Lebesgue point for any given  $w$  [17, Section 1.7]. We observe that every  $p$ -Lebesgue point is a  $q$ -Lebesgue point if  $1 \leq q \leq p$ ; if  $w \in L^\infty_{loc}(\Omega, \mu)^n$ , the notion does not depend on  $p$ . We will use the notion of Lebesgue points in particular for functions in the space  $L^p(\Gamma)^n$ , with  $\Gamma$  a  $C^1$  graph contained in  $\Omega$ —we note that this space coincides with  $L^p(\Omega, \mathcal{H}^{m-1} \llcorner \Gamma)^n$ .

## 2.3 Functions of bounded variation

Throughout the paper, we will consider convex functionals  $\mathcal{E}$  defined in  $L^1_{loc}(\Omega)^n$ ,  $n \geq 1$ , for  $\Omega \subset \mathbb{R}^m$  an open set. We will work with minimizers of  $\mathcal{E}$ , which will be assumed to belong to  $BV_{loc}(\Omega)^n$ . We recall that

$$BV(\Omega)^n = \left\{ w \in L^1(\Omega)^n : TV(w) < \infty \right\}$$

where the total variation  $TV$  is defined by

$$TV(w) = \sup \left\{ - \int w \operatorname{div} \varphi \, dx : \varphi \in C_c^\infty(\Omega; \mathbb{R}^m), |\varphi(x)| \leq 1 \text{ for } x \in \Omega \right\}.$$

It is easily checked (from Riesz's theorem) that  $TV(w)$  is finite if and only if the distributional derivative  $Dw$  is a bounded Radon measure in  $\Omega$ , in which case

$$TV(w) = \int_{\Omega} |Dw| = |Dw|(\Omega).$$

Then, one defines  $BV_{loc}(\Omega)^n = \bigcap_{A \subset \subset \Omega} BV(A)^n$ , where the intersection is on all open sets whose closure lies in  $\Omega$ .

By the Federer–Vol'pert theorem [3, Theorem 3.78], if  $w \in BV_{loc}(\Omega)^n$ , the set  $S_w$  is countably  $\mathcal{H}^{m-1}$ -rectifiable and  $\mathcal{H}^{m-1}(S_w \setminus J_w) = 0$ . In particular, the (multi)function  $\{w^+, w^-\}$  is defined  $\mathcal{H}^{m-1}$ -a. e. in  $\Omega$ . Thus also the precise representative  $\tilde{w}$  of  $w$  given by

$$\tilde{w} = (w^+ + w^-)/2$$

is defined up to  $\mathcal{H}^{m-1}$ -null sets. In general,  $w \in BV_{loc}(\Omega)^n$  admits one-sided traces on any oriented, countably  $\mathcal{H}^{m-1}$ -rectifiable subset of  $\Omega$ , see [3, Theorem 3.77]. Those traces coincide with  $w^\pm$   $\mathcal{H}^{m-1}$ -a. e. (up to permutation) [3, Remark 3.79].

The Radon measure  $Dw$  can be decomposed as:

$$Dw = D^a w + D^s w, \quad D^a w = \nabla w \mathcal{L}^m, \quad D^s w = D^c w + (w^+ - w^-) \otimes \nu_w \mathcal{H}^{m-1} \llcorner J_w$$

where

- $D^a w$  is the absolutely continuous part of  $Dw$ ,  $D^s w$  the singular part, and  $\nabla w \in L^1(\Omega)^{n \times m}$  is the Radon–Nikodym derivative of  $Dw$  with respect to the Lebesgue measure  $\mathcal{L}^m$ ;
- $D^c w$  is the ‘‘Cantor part’’ of  $Dw$ , which is singular with respect to the Lebesgue measure and vanishes on sets of finite  $(m-1)$ -dimensional Hausdorff measure  $\mathcal{H}^{m-1}$ ;
- the last term  $(w^+ - w^-) \otimes \nu_w \mathcal{H}^{m-1} \llcorner J_w$  is called the ‘‘jump part’’ of  $Dw$ .

The matrix  $(D^c w / |D^c w|)(x)$  appearing in the polar decomposition of the Cantor part  $D^c w = (D^c w / |D^c w|) |D^c w|$  is known to have rank one for  $|D^c w|$ -a. e.  $x \in \Omega$  [1], analogously to the jump part. We refer to [3] for more details.

Similarly [37],  $BD(\Omega)$  is defined as the space of displacements  $u \in L^1(\Omega)^m$  such that the symmetrized gradient  $Eu := (Du + Du^T)/2$  is a bounded Radon measure, and one has:

$$Eu = E^a w + E^s w, \quad E^a w = e(w) \mathcal{L}^m, \quad E^s w = E^c w + (w^+ - w^-) \odot \nu_w \mathcal{H}^{m-1} \llcorner J_w$$

with  $e(w) \in L^1(\Omega)^{m \times m}$  the approximate symmetrized gradient,  $E^c w$  the Cantor part and  $\odot$  the symmetrized tensor product  $(a \odot b := ((a_i b_j + a_j b_i)/2)_{i,j=1}^m)$ . Note that an analog of Alberti's rank one theorem also holds in  $BD$ , see [12].

### 3 Setting and main result

Let  $\Omega \subset \mathbb{R}^m$ ,  $m \geq 1$  be an open set. We consider functionals  $\mathcal{E}: L^1_{loc}(\Omega)^n \rightarrow [0, \infty]$  of form

$$\mathcal{E}(w) = \mathcal{F}(w - f) + \mathcal{R}(w),$$

where  $f \in L^1_{loc}(\Omega)^n$ . We assume the *fidelity*  $\mathcal{F}: L^1_{loc}(\Omega)^n \rightarrow [0, \infty]$  is given by

$$\mathcal{F}(w) = \int_{\Omega} \psi(w)$$

where  $\psi: \mathbb{R}^n \rightarrow [0, \infty[$  is convex. As for the *regularizer*  $\mathcal{R}: L^1_{loc}(\Omega)^n \rightarrow [0, \infty]$ , in general we only assume that it is convex, without prescribing a particular structure. The regularizer contains prior information of the reconstructed image  $u$ , and will usually be defined as a convex integral of the distributional gradient  $Dw$ , possibly with an additional box constraint  $w(x) \in K$  a. e. for some closed convex set  $K \subset \mathbb{R}^n$ , enforced by prescribing  $\mathcal{R}(w) = \infty$  if  $w$  does not satisfy it.

Our aim in this paper is to provide an estimate on the jumps of minimizers of  $\mathcal{E}$ , that is, functions  $u \in L^1_{loc}(\Omega)^n$  satisfying

$$\mathcal{E}(u) = \inf \left\{ \mathcal{E}(w) : w \in L^1_{loc}(\Omega)^n \right\}. \quad (5)$$

In the case that  $\psi$  is strictly convex, there is at most one  $u$ . However, without further assumptions,  $u$  might not exist.

Let  $\varphi \in C_c^\infty(\Omega)^n$ . For  $w \in L^1_{loc}(\Omega)^n$  and  $\tau \in \mathbb{R}$  with  $|\tau|$  sufficiently small we put

$$w_\tau^\varphi(x) = w(x + \tau\varphi(x)).$$

Suppose that  $\mathcal{R}(w) < \infty$ . We say that  $\mathcal{R}$  is *differentiable along inner variations* at  $w$  if the limit

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} (\mathcal{R}(w_\tau^\varphi) - \mathcal{R}(w)) \quad (6)$$

exists for all  $\varphi \in C_c^\infty(\Omega)^n$ . In practice, we will only use directional inner variations, where  $\varphi$  has the form  $\tilde{\varphi}\nu$  for  $\nu \in \mathbb{S}^{m-1}$  and  $\tilde{\varphi} \in C_c^\infty(\Omega)$ .

**Theorem 1.** *Suppose that  $f \in BV_{loc}(\Omega)^n$  and  $u$  is minimizing in (5) with  $\mathcal{E}(u) \neq \infty$ . We assume*

(H1)  $u \in BV_{loc}(\Omega)^n$ ,

(H2)  $u, f \in L^\infty_{loc}(\Omega)^n$  or  $D\psi$  is bounded,

(H3)  $\mathcal{R}$  is differentiable along inner variations at  $u$ .

If  $\psi$  is  $C^1$  and strictly convex, then  $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$ . If  $\psi$  is  $C^2$ , then

$$(u^+ - u^-) \cdot A(u^+ - u^-) \leq (f^+ - f^-) \cdot A(u^+ - u^-) \quad \mathcal{H}^{m-1}\text{-a. e. on } J_u, \quad (7)$$

where

$$A = \int_0^1 D^2\psi(u^- - f^- + s(u^+ - f^+ - u^- + f^-))ds.$$

In (7), the selections of  $u^\pm$  and  $f^\pm$  are chosen in a mutually consistent manner. Technically they are determined by a chosen orientation of the sequence of  $C^1$  graphs covering  $J_u$  (see Section 2.3), but evidently (7) does not depend on this choice. It follows from (7) that

$$(u^+ - u^-) \cdot A(u^+ - u^-) \leq (f^+ - f^-) \cdot A(f^+ - f^-),$$

which can be translated into a bound on the size of jumps of  $u$  in terms of  $f$ . In particular in the strongly convex, Lipschitz-gradient case

$$\lambda I \leq D^2\psi \leq \Lambda I \quad \text{with } 0 < \lambda \leq \Lambda, \quad (8)$$

we obtain

$$|u^+ - u^-| \leq \sqrt{\Lambda/\lambda} |f^+ - f^-|.$$

However, (7) also carries information about the jump direction in the value space  $\mathbb{R}^n$ . The proof of Theorem 1 is postponed after the proofs of the following two lemmata.

**Lemma 2.** *Let  $u$  be the minimizer of  $\mathcal{E}$ . Suppose that the limit (6) exists for  $w = u$ . For  $\vartheta \in [0, 1]$ , we denote*

$$u_{\vartheta, \tau}^\varphi = \vartheta u_\tau^\varphi + (1 - \vartheta)u.$$

Then, for  $\vartheta \in [0, 1]$ ,

$$\liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, \tau}^\varphi - f) - \mathcal{F}(u - f)) + \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, -\tau}^\varphi - f) - \mathcal{F}(u - f)) \geq 0. \quad (9)$$

*Proof.* By minimality,

$$0 \leq \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{E}(u_{\vartheta, \pm\tau}^\varphi) - \mathcal{E}(u)) = \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{R}(u_{\vartheta, \pm\tau}^\varphi) - \mathcal{R}(u) + \mathcal{F}(u_{\vartheta, \pm\tau}^\varphi - f) - \mathcal{F}(u - f)).$$

By our assumption, the function  $R_\varphi: ]-\tau_0, \tau_0[ \rightarrow [0, \infty[$  defined by

$$R_\varphi(\tau) = \mathcal{R}(u_\tau^\varphi)$$

for  $\tau_0$  small enough is differentiable at  $\tau = 0$ . Thus, by convexity of  $\mathcal{R}$ ,

$$\frac{1}{\tau} (\mathcal{R}(u_{\vartheta, \pm\tau}^\varphi) - \mathcal{R}(u)) \leq \frac{\vartheta}{\tau} (\mathcal{R}(u_{\pm\tau}^\varphi) - \mathcal{R}(u)) \rightarrow \pm\vartheta R'_\varphi(0) \quad \text{as } \tau \rightarrow 0.$$

Therefore,

$$0 \leq \pm\vartheta R'_\varphi(0) + \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, \pm\tau}^\varphi - f) - \mathcal{F}(u - f)).$$

We conclude by summing together the two obtained inequalities.  $\square$

We note that the Lemma 2 is the only place in the proof of Theorem 1 where we use convexity of  $\mathcal{R}$  (or differentiability of  $\mathcal{R}$  for that matter). In what follows, we will work directly with inequality (9). Thus, we could drop the convexity hypothesis altogether and instead assume explicitly that  $\mathcal{R}$  is differentiable along *mixed variations*  $u_{\vartheta, \tau}^\varphi$  (with fixed  $\vartheta \in [0, 1]$ ), leading directly to (9). In this way it might be possible to treat lower order perturbations of convex regularizers or the case of quasiconvex integrands, etc. However, we do not know simple and natural examples for which it is clear that such differentiability holds—one may check in particular that the celebrated Mumford–Shah functional [33] is not differentiable along mixed variations near jump points. In fact, also our proof of differentiability for concrete regularizers (Theorem 4) uses the duality formula for convex integrands. Thus, we decided to keep the simpler assumption of differentiability along inner variations in the statement of our main result and leave the discussion of non-convex regularizers to a possible future work.



**Lemma 3.** *Assume (H1) and (H2) hold. Let  $\Gamma \subset Q_{r_0}(x_0, \nu_0) \subset \Omega$  be a  $C^1$  graph admitting a graphical parametrization  $\gamma: B_{r_0}^{m-1}(x_0, \nu_0) \rightarrow Q_{r_0}(x_0, \nu_0)$ . Let  $\varphi \in C_c^\infty(\Omega)^m$  be such that the support of  $\varphi$  is contained in  $Q_r(x_0, \nu_0)$ ,  $0 < r \leq r_0$ , and  $\varphi = \nu_0 \tilde{\varphi}$ ,  $\tilde{\varphi} \in C_c^\infty(\Omega)$ . Moreover, assume that  $0 \leq \tilde{\varphi} \leq 1$ , restrictions of  $\tilde{\varphi}$  to lines parallel to  $\nu_0$  attain their maxima on  $\Gamma$ , and*

$$\frac{\partial \tilde{\varphi}}{\partial \nu_0}(x) = 0 \quad \text{for } x \text{ close to } \Gamma. \quad (10)$$

Then

$$\begin{aligned} & \limsup_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, \tau}^\varphi - f) - \mathcal{F}(u - f)) \\ & \leq \int_{\Gamma} \tilde{\varphi} \left( \psi(\vartheta u^+ + (1 - \vartheta)u^- - f^-) - \psi(u^- - f^-) \right) d\gamma_{\#} \mathcal{L}^{m-1} + C\vartheta |\nu_0 \cdot Du| (Q_r(x_0, \nu_0) \setminus \Gamma), \end{aligned}$$

$$\begin{aligned} & \limsup_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, -\tau}^\varphi - f) - \mathcal{F}(u - f)) \\ & \leq \int_{\Gamma} \tilde{\varphi} \left( \psi(\vartheta u^- + (1 - \vartheta)u^+ - f^+) - \psi(u^+ - f^+) \right) d\gamma_{\#} \mathcal{L}^{m-1} + C\vartheta |\nu_0 \cdot Du| (Q_r(x_0, \nu_0) \setminus \Gamma), \end{aligned}$$

where  $\gamma_{\#} \mathcal{L}^{m-1}$  denotes the pushforward of  $\mathcal{L}^{m-1}$  by  $\gamma$  and

$$C := \sup \left\{ |D\psi(\xi)| : \xi \in \mathbb{R}^n, |\xi| \leq \|u\|_{L^\infty(Q_{r_0}(x_0, \nu_0))^n} + \|f\|_{L^\infty(Q_{r_0}(x_0, \nu_0))^n} \right\} \quad (11)$$

is finite by virtue of (H2).

*Proof.* By an isometric change of coordinates, we will assume that  $\nu_0 = e_m$ ,  $x_0 = 0$  and denote  $x = (x', x_m)$ ,  $Q_r(x_0, \nu_0) = Q_r$ ,  $B_r^{m-1}(x_0, \nu_0) = B_r^{m-1}$ ,  $\gamma(x') = (x', \tilde{\gamma}(x'))$  for  $x' \in B_r^{m-1}$ , so that  $\Gamma = \{\gamma(x') : x' \in B_r^{m-1}\}$ . By our assumption, we have

$$x + \tau\varphi(x) = (x', x_m + \tau\tilde{\varphi}(x)).$$

We take  $\tau > 0$  small enough so that the map  $x \mapsto x + \tau\varphi(x)$  is a diffeomorphism. In particular,  $x_m \mapsto x_m + \tau\tilde{\varphi}(x', x_m)$  is a diffeomorphism for every  $x' \in B_r^{m-1}$ . By (10), we can also assume

$$\tilde{\varphi}(x', x_m) = \tilde{\varphi}(\gamma(x')) \quad \text{whenever} \quad -\tau\tilde{\varphi}(x', x_m) \leq x_m - \tilde{\gamma}(x') \leq \tau\tilde{\varphi}(x', x_m).$$

We will prove the first part of the assertion. The proof of the second one is the same. We rewrite

$$\begin{aligned} \mathcal{F}(u_{\vartheta, \tau}^\varphi - f) - \mathcal{F}(u - f) &= \int_{B_r^{m-1}} \int_{\tilde{\gamma}(x') - \tau\tilde{\varphi}(x')}^{\tilde{\gamma}(x')} \psi(u_{\vartheta, \tau}^\varphi - f) - \psi(u - f) dx_m dx' \\ &+ \int_{B_r^{m-1}} \int_{-r, r \setminus [\tilde{\gamma}(x') - \tau\tilde{\varphi}(\gamma(x')), \tilde{\gamma}(x')]} \psi(u_{\vartheta, \tau}^\varphi - f) - \psi(u - f) dx_m dx'. \end{aligned}$$

Using [3, Theorem 3.108], for  $\mathcal{L}^{m-1}$ -a. e.  $x' \in B_r^{m-1}$  we have

$$\frac{1}{\tau} \int_{\tilde{\gamma}(x') - \tau\tilde{\varphi}(\gamma(x'))}^{\tilde{\gamma}(x')} \psi(u - f) dx_m \rightarrow \tilde{\varphi} \psi(u^- - f^-) \Big|_{\gamma(x')},$$

$$\frac{1}{\tau} \int_{\tilde{\gamma}(x') - \tau \tilde{\varphi}(\gamma(x'))}^{\tilde{\gamma}(x')} \psi(u_{\vartheta, \tau}^{\varphi} - f) dx_m \rightarrow \tilde{\varphi} \psi(\vartheta u^+ + (1 - \vartheta)u^- - f^-) \Big|_{\gamma(x')},$$

where we recall that we have chosen  $u^-$ ,  $f^-$  (resp.  $u^+$ ,  $f^+$ ) to be the approximate limits corresponding to traces of  $u$ ,  $f$  along  $\Gamma$  “from below” (resp. “from above”), consistently with the choice of  $u^\pm(x_0)$  given by  $\nu_0$ .

On the other hand, identifying  $u$  with its precise representative, the slicing properties of  $BV$  functions [3, §3.11, Theorem 3.107] ensure that for  $\mathcal{L}^{m-1}$ -a. e.  $x' \in B_r^{m-1}$  the function  $u_{x'}: x_m \mapsto u(x', x_m)$  is in  $BV([-r, r])^n$ , and we can write for  $\mathcal{L}^1$ -a. e.  $x_m \in ]-r, r[$ :

$$u_{\vartheta, \tau}^{\varphi}(x', x_m) - u(x', x_m) = \vartheta Du_{x'} (]x_m, x_m + \tau \tilde{\varphi}(x', x_m)[).$$

Therefore

$$\begin{aligned} \left| \psi(u_{\vartheta, \tau}^{\varphi} - f) - \psi(u - f) \right| (x', x_m) &= \int_0^1 D\psi(u + t(u_{\vartheta, \tau}^{\varphi} - u) - f) dt \cdot (u_{\vartheta, \tau}^{\varphi} - u) \\ &\leq C\vartheta |Du_{x'}| (]x_m, x_m + \tau \tilde{\varphi}(x', x_m)[) \\ &\leq C\vartheta |Du_{x'}| (]x_m, x_m + \tau \tilde{\varphi}(\gamma(x'))[), \end{aligned}$$

where  $C$  is defined in (11) and we have used that  $\tilde{\varphi}(x', \cdot)$  is maximal at  $\tilde{\gamma}(x')$ . Thus, for  $\mathcal{L}^{m-1}$ -a. e.  $x' \in B_r^{m-1}$  and small enough  $\tau > 0$ ,

$$\begin{aligned} \frac{1}{\tau} \int_{]-r, r[ \setminus [\tilde{\gamma}(x') - \tau \tilde{\varphi}(\gamma(x')), \tilde{\gamma}(x')]} \left| \psi(u_{\vartheta, \tau}^{\varphi} - f) - \psi(u - f) \right| dx_m \\ \leq C \frac{\vartheta}{\tau} \int \int \chi_{]-r, r[ \setminus [\tilde{\gamma}(x') - \tau \tilde{\varphi}(\gamma(x')), \tilde{\gamma}(x')]}(x_m) \chi_{]x_m, x_m + \tau \tilde{\varphi}(\gamma(x'))[}(t) d|Du_{x'}|(t) dx_m. \end{aligned}$$

The integrand is not zero only when  $x_m < t < x_m + \tau \tilde{\varphi}(\gamma(x'))$  and either  $-r < x_m < \tilde{\gamma}(x') - \tau \tilde{\varphi}(\gamma(x'))$  or  $\tilde{\gamma}(x') < x_m \leq r$ , in particular one has  $t \in ]-r, r[ \setminus \{\tilde{\gamma}(x')\}$  and  $t - \tau \tilde{\varphi}(\gamma(x')) \leq x_m \leq t$ . Using Fubini's theorem, we deduce that this expression is bounded by

$$C\vartheta \tilde{\varphi}(\gamma(x')) |Du_{x'}| (]-r, r[ \setminus \{\tilde{\gamma}(x')\}).$$

Appealing to [3, Theorem 3.107],

$$\frac{1}{\tau} \int_{B_r^{m-1}} \int_{]-r, r[ \setminus [\tilde{\gamma}(x') - \tau \tilde{\varphi}(\gamma(x')), \tilde{\gamma}(x')]} \psi(u_{\vartheta, \tau}^{\varphi} - f) - \psi(u - f) dx_m dx' \leq C\vartheta |\nu_0 \cdot Du|(Q_r \setminus \Gamma).$$

□

*Proof of Theorem 1.* Let  $(\Gamma_i)_{i=1}^\infty$  be a sequence of  $C^1$  graphs that covers  $J_u$  up to a  $\mathcal{H}^{m-1}$ -null set. Let us fix an index  $i$ . By [3, eq. (2.41) on p. 79], for  $\mathcal{H}^{m-1}$ -a. e.  $x_0$  in  $J_u \cap \Gamma_i$

$$\Theta^{m-1}(|Du| \llcorner (\Omega \setminus \Gamma_i), x_0) = 0. \quad (12)$$

We choose such an  $x_0$ , and assume in addition that  $x_0$  is a Lebesgue point for  $u^\pm$  and  $f^\pm$  with respect to  $\mathcal{H}^{m-1} \llcorner \Gamma_i$ . We take  $\nu_0 = \nu_u(x_0)$ . For  $r_0 > 0$  small enough  $Q_{r_0}(x_0, \nu_0) \subset \Omega$  and  $\Gamma := \Gamma_i \cap Q_{r_0}(x_0, \nu_0)$  has a graphical parametrization  $\gamma: B_{r_0}^{m-1}(x_0, \nu_0) \rightarrow Q_{r_0}(x_0, \nu_0)$ . Moreover, since  $\Gamma$  is tangent to  $x_0 + \nu_0^\perp$  at  $x_0$ , possibly decreasing  $r_0$ , we can assume that

$\gamma(x') \in x' + [-r/2, r/2]\nu_0$  for  $x' \in B_r^{m-1}(x_0, \nu_0)$ ,  $0 < r < r_0$  and construct a sequence of  $\tilde{\varphi}$  satisfying the assumptions of Lemma 3 that converges to 1 on  $Q_r(x_0, \nu_0) \cap \Gamma$ . Then, by Lemmata 2 and 3,

$$0 \leq \int_{Q_r(x_0, \nu_0) \cap \Gamma} \psi(\vartheta u^+ + (1 - \vartheta)u^- - f^-) - \psi(u^- - f^-) d\gamma_{\#} \mathcal{L}^{m-1} \\ + \int_{Q_r(x_0, \nu_0) \cap \Gamma} \psi(\vartheta u^- + (1 - \vartheta)u^+ - f^+) - \psi(u^+ - f^+) d\gamma_{\#} \mathcal{L}^{m-1} + 2C\vartheta |\nu_0 \cdot Du| (Q_r(x_0, \nu_0) \setminus \Gamma).$$

Dividing by  $\mathcal{L}^{m-1}(B_r^{m-1}) \sim r^{m-1}$  and passing to the limit  $r \rightarrow 0^+$ , we get

$$0 \leq \psi(\vartheta u^+ + (1 - \vartheta)u^- - f^-) - \psi(u^- - f^-) + \psi(\vartheta u^- + (1 - \vartheta)u^+ - f^+) - \psi(u^+ - f^+)$$

at  $x_0$  by (12). By convexity of  $\psi$ ,

$$\psi(\vartheta u^+ + (1 - \vartheta)u^- - f^-) - \psi(u^- - f^-) \leq \vartheta D\psi(\vartheta u^+ + (1 - \vartheta)u^- - f^-) \cdot (u^+ - u^-),$$

$$\psi(\vartheta u^- + (1 - \vartheta)u^+ - f^+) - \psi(u^+ - f^+) \leq \vartheta D\psi(\vartheta u^- + (1 - \vartheta)u^+ - f^+) \cdot (u^- - u^+).$$

Summing up,

$$0 \leq \vartheta \left( D\psi(\vartheta u^+ + (1 - \vartheta)u^- - f^-) - D\psi(\vartheta u^- + (1 - \vartheta)u^+ - f^+) \right) \cdot (u^+ - u^-).$$

Dividing the inequality by  $\vartheta$  and letting  $\vartheta \rightarrow 0^+$  yields

$$0 \leq \left( D\psi(u^- - f^-) - D\psi(u^+ - f^+) \right) \cdot (u^+ - u^-). \quad (13)$$

If  $f^+ = f^-$  and  $\psi$  is strictly convex, then  $D\psi$  is strictly monotone and we get a contradiction unless  $u^+ = u^-$ : this shows that  $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$ . On the other hand, if  $\psi \in C^2$ , applying the fundamental theorem of calculus to the function

$$s \mapsto D\psi \left( u^- - f^- + s(u^+ - f^+ - u^- + f^-) \right) \cdot (u^+ - u^-),$$

and using (13) we obtain (7). □

*Remark 1.* Assumption (H1) will hold as soon as  $\mathcal{R}$  is coercive in  $BV(\Omega)^n$ ; the next section 4 discusses a class of such regularizers for which (H3) holds as well. As for (H2), it is trivially satisfied if  $\psi$  is globally Lipschitz. Otherwise, we need to ensure that  $u$  and  $f$  are locally bounded. Assuming that  $f \in L^\infty(\Omega)^n$ , a minimizer of  $\mathcal{E}$  will be bounded if  $n = 1$  and  $\psi$  is coercive, while if  $n \geq 1$  this requires various types of assumptions on  $\psi$  and  $\mathcal{R}$ , see Section 5. Alternatively, we can enforce it by considering a constrained minimization problem, which amounts to replacing  $\mathcal{R}$  with

$$\mathcal{R}_K := \mathcal{R} + \iota_{\{w \in L^1_{loc}(\Omega)^n : w(x) \in K \text{ for a.e. } x \in \Omega\}}$$

where  $K$  is bounded, closed and convex (here  $\iota_U$  denotes the convex-analytic characteristic function of  $U \subset L^1_{loc}(\Omega)^n$ , which is 0 in  $U$  and  $\infty$  outside). We observe that if  $\mathcal{R}$  satisfies (H3), then  $\mathcal{R}_K$  satisfies it as well. Moreover, if  $\mathcal{R}$  is coercive in  $BV(\Omega)^n$  (that is,  $w \in BV(\Omega)^n$  whenever  $\mathcal{R}(w) < \infty$ ), then  $\mathcal{R}_K$  is as well.

## 4 Differentiable regularizers

### 4.1 General form

In this section we give some examples of functionals  $\mathcal{R}$  whose domain contains  $BV(\Omega)^n$  and that are differentiable along inner variations, in particular assumptions (H1) and (H3) in Theorem 1 are satisfied whenever  $\mathcal{R}(u) < \infty$ . We assume for simplicity that  $\Omega$  is bounded. Let  $\varrho: \mathbb{R}^{n \times m} \rightarrow [0, \infty[$  be convex and satisfy

$$\varrho(A) \leq C(1 + |A|) \quad \text{for } \xi \in \mathbb{R}^{n \times m} \quad (14)$$

with  $C > 0$ . For a vector Radon measure  $\mu$  on  $\Omega$  we denote by  $\mu^a$  its density with respect to the Lebesgue measure on  $\Omega$  and by  $\mu^s$  its Lebesgue singular part, i. e.

$$d\mu = \mu^a d\mathcal{L}^m + d\mu^s, \quad \mathcal{L}^m \perp \mu^s.$$

Moreover, for  $\xi \in \mathbb{R}^{n \times m}$  we denote

$$\varrho^\infty(\xi) = \lim_{t \rightarrow \infty} \frac{1}{t} \varrho(t\xi).$$

Then

$$d\varrho(\mu) = \varrho(\mu^a) d\mathcal{L}^m + \varrho^\infty\left(\frac{\mu^s}{|\mu^s|}\right) d|\mu^s|$$

defines a Radon measure  $\varrho(\mu)$  and

$$\mathcal{R}(w) = \begin{cases} \varrho(Dw)(\Omega) & \text{if } w \in BV(\Omega)^n, \\ \infty & \text{otherwise} \end{cases} \quad (15)$$

defines a convex functional on  $L^1_{loc}(\Omega)$  [14]. At least when  $\varrho$  satisfies a lower bound  $\varrho(A) \geq C'|A|$  with  $C' > 0$ ,  $\mathcal{R}$  is lower semicontinuous and  $\mathcal{E}$  attains its minimum at  $u \in BV(\Omega)$ .

**Theorem 4.** *Suppose that*

(D1) *the function  $\tau \mapsto \varrho(A(I + \tau B))$  is differentiable at  $\tau = 0$  for any  $A \in \mathbb{R}^{n \times m}$  and any  $B \in \mathbb{R}^{m \times m}$ ,*

(D2) *the function  $\tau \mapsto \varrho^\infty(A(I + \tau B))$  is differentiable at  $\tau = 0$  for any  $A \in \mathbb{R}^{n \times m}$  of rank 1 and any  $B \in \mathbb{R}^{m \times m}$ .*

*Then  $\mathcal{R}$  is differentiable along inner variations at any  $w \in \mathcal{D}$ .*

The proof of Theorem 4 follows along the lines of [18, Chapter 10] (where the case of total variation is considered). The main point is the following change of variables formula.

**Lemma 5.** *Let  $F: \Omega \rightarrow \Omega$  be a diffeomorphism ( $C^1$  up to the boundary) and let  $w \in BV(\Omega)^n$ . Then*

$$\mathcal{R}(w \circ F) = \int_{\Omega} |\det D(F^{-1})| d\varrho(Dw DF \circ F^{-1}). \quad (16)$$

*Proof.* We will use the dual representation of  $\mathcal{R}$  as given in [14, Lemma 1.1], according to which

$$\int_{\Omega} \varphi d\varrho(\mu) = \sup_{h \in D_{\varrho}} \int_{\Omega} \varphi h \cdot d\mu - \int_{\Omega} \varphi \varrho^*(h) d\mathcal{L}^m, \quad (17)$$

for  $\mu \in M(\Omega)^{n \times n}$  and  $\varphi \in C(\overline{\Omega})$  with  $\varphi \geq 0$ , where

$$D_\varrho = \{h \in C_c(\Omega)^{n \times m} : \varrho^*(h) \in L^1(\Omega)\}.$$

We recall [18, Thm. 1.17][3, Thm. 3.9] that for each  $w \in BV(\Omega)^n$  there exists a sequence  $(w_k) \subset C^\infty(\Omega)^n$  such that  $w_k \rightarrow w$  in  $L^1(\Omega)^n$  and

$$\int_{\Omega} |Dw_k| d\mathcal{L}^m \rightarrow |Dw|(\Omega),$$

in particular  $Dw_k \xrightarrow{*} Dw$  in  $M(\Omega)^{n \times m}$ . It is easy to check that  $w_k \circ F \rightarrow w \circ F$  in  $L^1(\Omega)^n$ . We also have

$$\int_{\Omega} |D(w_k \circ F)| \leq \int_{\Omega} |Dw_k \circ F| |DF| = \int_{\Omega} |Dw_k| |DF \circ F^{-1}| |\det D(F^{-1})|,$$

in particular  $w_k \circ F$  is bounded in  $BV(\Omega)^n$ , whence  $w \circ F \in BV(\Omega)^n$  and

$$D(w_k \circ F) \xrightarrow{*} D(w \circ F) \quad \text{in } M(\Omega)^{n \times m}. \quad (18)$$

For any  $h \in C_c(\Omega)^{n \times m}$ , we can calculate

$$\int_{\Omega} h \circ F \cdot D(w_k \circ F) = \int_{\Omega} h \circ F \cdot Dw_k \circ F DF = \int_{\Omega} h \cdot Dw_k DF \circ F^{-1} |\det D(F^{-1})|.$$

Using (18), we pass to the limit  $k \rightarrow \infty$  obtaining

$$\int_{\Omega} h \circ F \cdot dD(w \circ F) = \int_{\Omega} |\det D(F^{-1})| h \cdot dDw DF \circ F^{-1}. \quad (19)$$

Note that

$$\int_{\Omega} \varrho^*(h \circ F) = \int_{\Omega} |\det D(F^{-1})| \varrho^*(h), \quad (20)$$

in particular  $h \in D_\varrho$  iff  $h \circ F \in D_\varrho$ . Thus, using the dual representation formula (17) in conjunction with (19) and (20), we obtain the desired equality.  $\square$

*Proof of Theorem 4.* We take  $F_\tau(x) = x + \tau\varphi(x)$  with  $\varphi \in C_c^\infty(\Omega)^m$ ,  $x \in \Omega$  and  $\tau$  small enough. By the representation formula (16) and recalling the definition of the measure  $\varrho(Dw DF_\tau \circ F_\tau^{-1})$ ,

$$\begin{aligned} \mathcal{R}(w \circ F_\tau) &= \int_{\Omega} \det D(F_\tau^{-1}) \varrho(Dw^a DF_\tau \circ F_\tau^{-1}) d\mathcal{L}^m \\ &\quad + \int_{\Omega} \det D(F_\tau^{-1}) \varrho^\infty \left( \frac{Dw^s}{|Dw^s|} DF_\tau \circ F_\tau^{-1} \right) d|Dw^s|. \end{aligned} \quad (21)$$

Note that  $DF_\tau \circ F_\tau^{-1} = I + \tau D\varphi \circ F_\tau^{-1}$ . Denoting  $\varrho_x(\tau) = \varrho(Dw^a(x) DF_\tau \circ F_\tau^{-1}(x))$ , we have for  $\tau \neq 0$

$$\begin{aligned} \frac{1}{\tau}(\varrho_x(\tau) - \varrho_x(0)) &= \frac{1}{\tau}(\varrho(Dw^a(x)(I + \tau D\varphi(x))) - \varrho(Dw^a(x))) \\ &\quad + \frac{1}{\tau}(\varrho(Dw^a(x)(I + \tau D\varphi \circ F_\tau^{-1}(x))) - \varrho(Dw^a(x)(I + \tau D\varphi(x)))) \end{aligned}$$

Using assumption (D1), the global Lipschitz continuity of  $\varrho$  and the continuity of  $\tau \mapsto F_\tau^{-1}$ , we deduce that  $\varrho_x$  is differentiable at  $\tau = 0$  for  $\mathcal{L}^m$ -a. e.  $x \in \Omega$  and there exists  $C_{\varrho,\varphi} > 0$  such that

$$\frac{1}{\tau}|\varrho_x(\tau) - \varrho_x(0)| \leq C_{\varrho,\varphi}|Dw^a(x)|.$$

Thus, by dominated convergence, the first integral in (21) is differentiable at  $\tau = 0$ . Similarly, appealing to Alberti's rank-one theorem [3, Thm. 3.94] [1] and assumption (D2), we show that the second integral in (21) is differentiable at  $\tau = 0$ .  $\square$

## 4.2 Examples

A simple way to ensure that conditions (D1) and (D2) of Theorem 4 are satisfied is to assume that  $\varrho$  and  $\varrho^\infty$  are differentiable everywhere except at 0. In particular, *vectorial* (such as defined by the Frobenius norm) and *anisotropic (vectorial) total variations* given by any norm on  $\mathbb{R}^{n \times m}$  that is differentiable outside 0, such as  $\ell^{p,q}$ ,  $p, q \in ]1, \infty[$ , are differentiable along inner variations. The same is not true in the limiting cases  $1, \infty$ . In fact there are known examples where the assertion of Theorem 1 fails in these cases, see [26].

A more striking example is given by the Nuclear or Trace norm:

$$\varrho(A) = \text{Trace}((AA^T)^{\frac{1}{2}}), \quad (22)$$

which is the sum of the singular values. The function:

$$\tau \mapsto A(I + \tau B)(I + \tau B)^T A^T$$

is a one-parameter analytic function with values symmetric  $n \times n$  matrices, so that by [35, p. 39], the squared singular values can also be described by analytic functions (and in particular, in a neighborhood of  $\tau = 0$ , the positive eigenvalues stay positive). In addition, the kernel of  $(I + \tau B)^T A^T$  is the same as the kernel of  $A^T$  for small  $\tau$ , so that the number of non-zero singular values remains constant near 0. We deduce that the sum of the singular values is also an analytic function near 0 so that  $\varrho$  satisfies the assumptions of Theorem 4.

In addition, we also deduce that any convex and one-homogeneous, differentiable and symmetric function of the singular values will enjoy the same properties, such as the  $p$ -Schatten norms for  $p \in ]1, \infty[$ . (The  $p$ -Schatten norm is the  $\ell_p$  norm of the singular values.)

More generally for  $A \in \mathbb{R}^{n \times m}$ , letting  $p = \min\{m, n\}$  we denote  $\sigma(A) = (\sigma_1(A), \dots, \sigma_p(A))$  the ordered singular values of  $A$ . We give a simple proof of the following standard result on "unitary invariant" convex functions of matrices, as a corollary of von Neumann's inequality.

**Proposition 6.** *Let  $\tilde{h}$  be a proper, lower semicontinuous, extended real-valued convex function on  $\mathbb{R}_+^p$ , non-decreasing with respect to each coordinate. Then*

$$h(A) := \tilde{h}(\sigma(A)) = \tilde{h}(\sigma_1(A), \dots, \sigma_p(A))$$

*defines a convex function on  $\mathbb{R}^{n \times m}$ .*

*Proof.* We first observe that if we extend  $\tilde{h}$  to the whole  $\mathbb{R}^p$  as an even function with respect to each coordinate:

$$\tilde{h}(s_1, \dots, s_p) = \tilde{h}(|s_1|, \dots, |s_p|) \quad \text{for } (s_1, \dots, s_p) \in \mathbb{R}^p, \quad (23)$$

then  $\tilde{h}$  is convex on  $\mathbb{R}^p$ . For any extended-real valued function on  $\mathbb{R}^p$  satisfying (23), the same holds for its convex conjugate (and biconjugate). Moreover, if  $t \in \mathbb{R}_+^p$ , then

$$\tilde{h}^*(t) = \sup_{s \in \mathbb{R}_+^p} t \cdot s - \tilde{h}(s).$$

It then follows from von Neumann's inequality [34, 32]  $\text{Tr } AB^T \leq \sum_{i=1}^p \sigma_i(A)\sigma_i(B)$  that

$$h^*(B) = \sup_{A \in \mathbb{R}^{n \times m}} \text{Tr } AB^T - h(A) \leq \sup_{s \in \mathbb{R}_+^p} s \cdot \sigma(B) - \tilde{h}(s) = \tilde{h}^*(\sigma(B)),$$

while the opposite inequality is obvious, choosing  $A$  a matrix with singular values  $s \in \mathbb{R}_+^p$  and the same singular vectors as  $B$ . By the same reasoning and the Fenchel-Moreau theorem, we deduce

$$h^{**}(A) = \sup_{B \in \mathbb{R}^{m \times n}} \text{Tr } AB^T - h^*(B) = \sup_{s \in \mathbb{R}_+^p} \sigma(A) \cdot s - \tilde{h}^*(s) = \tilde{h}^{**}(\sigma(A)) = \tilde{h}(\sigma(A)) = h(A).$$

This shows that  $h$  is convex. □

We remark that if in addition  $\tilde{h}$  is smooth and a symmetric function of its arguments, then the discussion above for the Nuclear norm applies and  $h$  is differentiable along inner variations. An interesting example is the following: we consider

$$\varrho(A) = \log \sum_{i=1}^p \exp(\sigma_i(A)).$$

We claim that  $\varrho$  satisfies the assumptions of Theorem 4. Indeed, on the one hand,  $\varrho(A)$  is smooth and satisfies (D1). On the other hand, one readily checks that

$$\varrho^\infty(A) = \lim_{t \rightarrow +\infty} \frac{1}{t} \varrho(tA) = \max\{\sigma_1(A), \dots, \sigma_p(A)\}.$$

Therefore  $\varrho^\infty$  is the Spectral (or Operator) norm, which does not satisfy (D1), yet satisfies (D2) since it coincides with the Frobenius (as well as Nuclear) norm on rank-one matrices.

## 5 Boundedness of minimizers

In this section we consider minimizers of  $\mathcal{E}$  with  $\mathcal{R}(w) = \int_\Omega \varrho(Dw)$ ,  $\varrho$  satisfying the assumptions of Theorem 4 (whence (H3) holds), and  $\mathcal{F}(w) = \int_\Omega \psi(w - f)$  with  $\psi$  convex. In order to show that Theorem 1 applies, we need to check that (H1) and (H2) are also satisfied. We first assume that  $\varrho$  is coercive ( $\varrho(A) \geq c(|A| - 1)$  for some  $c > 0$ ), so that (H1) trivially holds. As for (H2), the situation is trivial if  $\psi$  is Lipschitz. Otherwise, as already mentioned in Remark 1, we can assume that  $f$  is bounded and ensure that the domain of  $\mathcal{R}$  is contained in  $L^\infty(\Omega)^n$  by imposing a box constraint.

That being said, let us now consider the case of unconstrained functional  $\mathcal{R}$  given by (15).

**Scalar case** The easiest is the scalar case  $n = 1$ .

**Lemma 7.** *Let  $n = 1$  and assume that  $\psi$  is coercive, that is  $\lim_{t \rightarrow \pm\infty} \psi(t) = \infty$ . Let  $f \in L^\infty(\Omega)$  and let  $u \in BV(\Omega)$  be a minimizer of  $\mathcal{E}$ . Then  $u \in L^\infty(\Omega)$ .*

*Proof.* By assumption, there is  $T > 0$  such that  $\psi$  is decreasing on  $] -\infty, T[$  and increasing on  $]T, \infty[$ . Let  $u^M := M \wedge (u \vee -M)$  (the function  $u \vee v$  is  $x \mapsto \max\{u(x), v(x)\}$ , and similarly  $u \wedge v$  is the minimum of  $u$  and  $v$ ), and let  $M > T + \|f\|_{L^\infty(\Omega)^n}$ . Suppose that  $|u| > M$  on a set  $E \subset \Omega$  of positive measure. For a.e.  $x \in E$ , if  $u(x) > M$ , then  $u(x) - f(x) > u^M(x) - f(x) > T$  and if  $u(x) < -M$ , then  $u(x) - f(x) < u^M(x) - f(x) < -T$ , whence  $\psi(u(x) - f(x)) > \psi(u^M(x) - f(x))$ . It follows that  $\mathcal{F}(u^M - f) < \mathcal{F}(u - f)$ .

On the other hand, it is well known that  $\mathcal{R}(u^M) < \mathcal{R}(u)$ , unless  $u^M = u$ : this can be deduced from the Chain Rule [2] which shows that:

$$\varrho(Du^M) = \varrho(D^a u) \chi_{\{|u| \leq M\}} + \varrho(D^c u) \chi_{\{|\tilde{u}| \leq M\}} + ((u^M)^+ - (u^M)^-) \varrho^\infty(\nu_u) \mathcal{H}^{m-1} \llcorner J_u,$$

and is strictly below  $\varrho(Du)$  if  $u^M \neq u$ . (Here,  $\tilde{u}$  denotes the precise representative, see Section 2.3.) It follows that  $\mathcal{E}(u^M) < \mathcal{E}(u)$ , a contradiction.  $\square$

**Vectorial case** The vectorial case is more complicated. A criterion for having a maximum principle in vectorial variational problems is identified in [29, 30]. Our criterion for the regularizer is derived from these references, and ensures that when  $u$  is projected on some half-space in some (at least  $n$ ) directions, then  $\mathcal{R}$  will decrease. For the data term, we need also that  $\mathcal{F}(u - f)$  decreases along certain projections, which is ensured for instance if  $\psi$  is uniformly coercive in such directions, in the sense which we propose below. We assume that there exist  $(e_1, \dots, e_n)$  independent unit vectors of  $\mathbb{R}^n$  such that for  $i = 1, \dots, n$ :

- (i)  $\varrho((I - e_i \otimes e_i)A) \leq \varrho(A)$  for all  $A \in \mathbb{R}^{n \times m}$  (with strict inequality if  $A^T e_i \neq 0$ );
- (ii) Letting, for  $u' \in e_i^\perp$ ,

$$t_i(u') := \sup \left\{ |t_*| : t_* \in \arg \min_{t \in \mathbb{R}} \psi(u' + te_i) \right\},$$

one has  $T_i := \sup_{u' \in e_i^\perp} t_i(u') < \infty$ .

When (ii) holds, we observe that  $t \mapsto \psi(u' + te_i)$  is increasing for  $t > T_i$  and decreasing for  $t < -T_i$  by convexity of  $\psi$ .

Thus, we can reproduce the proof of Lemma 7 in each direction  $e_i$ , using

$$u(x) - (e_i \cdot u(x) - M)_+ e_i - (e_i \cdot u(x) + M)_- e_i$$

in place of  $M \wedge (u \vee -M)$ . We obtain:

**Lemma 8.** *Assume that (i) and (ii) above hold. Let  $f \in L^\infty(\Omega)^n$  and let  $u \in BV(\Omega)^n$  be a minimizer of  $\mathcal{E}$ . Then  $u \in L^\infty(\Omega)^n$ .*

**Examples** For all the examples in Section 4.2, (i) is true in all directions of the canonical basis (and in all directions for the Frobenius norm, or the Schatten norms or other symmetric and non-decreasing function of the singular values, see [5, Prop. 6.4]). Hence Theorem 1 holds for minimizers of  $\mathcal{E}$ , for many data terms, such as data terms of the form  $\psi(w) = \tilde{\psi}(|w|)$  with  $\tilde{\psi}$  a non-negative convex function.



**A remark on the non coercive case** We give an (elementary) example here of a situation where the energy is not even coercive in  $BV(\Omega)^n$  and yet, Theorem 1 still applies. We consider  $\omega \subset \mathbb{R}^{m-1}$  and  $\Omega = \omega \times \mathbb{S}^1 \subset \mathbb{R}^m / (\{0_{m-1}\} \times \mathbb{Z})$  a  $m$ -dimensional open set which is periodic in the last variable  $x_m$ . For simplicity we set  $n = 1$ . We let  $\varrho(p) = \sqrt{\sum_{i=1}^{m-1} p_i^2}$ , for  $p \in \mathbb{R}^m$ , and to simplify  $\psi(t) = t^2/2$ . Then, let  $f \in BV(\Omega) \times L^\infty(\Omega)$  and  $u$  be the unique minimizer of

$$\varrho(Du)(\Omega) + \frac{1}{2} \int_{\Omega} (u - f)^2.$$

Observe that a minimizing sequence is bounded in  $L^2(\Omega)$  and will converge (weakly) to some limit. The regularizer being convex and lower semicontinuous, a minimizer exists (and is unique by strict convexity). A priori,  $u$  is not  $BV$ , as only the first  $(m - 1)$  components of  $Du$  are bounded measures. Yet, as the problem is also a one-parameter family of ‘‘ROF’’ type problems (1), it enjoys similar properties and in particular is non-expansive in any  $L^p(\Omega)$ ,  $1 \leq p \leq \infty$ . Hence, one has for any  $t \in \mathbb{R}$ , denoting  $e_m = (0, \dots, 0, 1)$ :

$$\int_{\Omega} |u(x + te_m) - u(x)| dx \leq \int_{\Omega} |f(x + te_m) - f(x)| dx \leq |t| \int_{\Omega} |Df|$$

and we deduce that  $e_m \cdot Du$  is also a bounded measure. As clearly one also has  $\|u\|_{\infty} \leq \|f\|_{\infty} < \infty$ , it follows that  $u$  satisfies the assumptions of Theorem 1 and we deduce  $u^+ - u^- \leq f^+ - f^-$   $\mathcal{H}^{m-1}$ -a. e. in  $\Omega$ .

## 6 Higher order regularizers

### 6.1 Inf-convolution based higher order regularizers

Here we consider regularizers of form

$$\mathcal{R}(w) = \inf_{z \in \mathcal{D}_2} \tilde{\mathcal{R}}(w, z). \quad (24)$$

where  $\mathcal{D}_2$  is, in general, a set and  $\tilde{\mathcal{R}}: L^1_{loc}(\Omega)^n \times \mathcal{D}_2 \rightarrow [0, \infty]$  is convex. This includes the following several variants of  $TV$  of inf-convolution type, introduced in literature to remedy the phenomenon of staircasing observed in solutions to (1). For simplicity, we recall their form in the case  $n = 1$ .

- *Total generalized variation (of second order)*

$$TGV(w) = \min_{z \in BD(\Omega)} |Dw - z^T \mathcal{L}^m|(\Omega) + |Ez|(\Omega),$$

where  $Ez = \frac{1}{2}(Dz + Dz^T)$  is the symmetrized gradient and

$$BD(\Omega) = \{z \in L^1(\Omega)^m : Ez \in M(\Omega)^{m \times m}\}$$

is the space of functions bounded deformation [37].

- *Non-symmetrized variant of TGV,*

$$nsTGV(w) = \min_{z \in BV(\Omega)^m} |Dw - z^T \mathcal{L}^m|(\Omega) + |Dz|(\Omega).$$

- *Infimal convolution total variation (of second order),*

$$ICTV(w) = \min_{z \in BV^2(\Omega)} |Dw - Dz|(\Omega) + |DDz|(\Omega),$$

where

$$BV^2(\Omega) = \{z \in BV(\Omega) : Dz \in BV(\Omega)^m\}.$$

We will produce a version of Theorem 1 that applies to smooth variants of all these examples. In our current setting, varying the whole functional  $\mathcal{R}$  in the direction of variable  $w$  is not a natural approach. Instead, we will use the formal equivalence of the minimization problem (5) for  $\mathcal{E}$  with the problem of finding  $u, v$  such that

$$\tilde{\mathcal{E}}(u, v) = \inf\{\mathcal{E}(w, z) : w \in L^1_{loc}(\Omega)^n, z \in \mathcal{D}_2\}, \quad \text{where } \tilde{\mathcal{E}}(w, z) := \tilde{\mathcal{R}}(w, z) + \mathcal{F}(w - f) \quad (25)$$

and consider suitable variations that move both  $w$  and  $z$ . Another issue, appearing for example in the case of  $TGV$ , is that  $\mathcal{D}_2$  might not be closed under (directed) inner variations. On the bright side, we do not need to assume any particular form of variation in direction of  $v$ , since we are not interested in obtaining bounds on the part of the minimizer corresponding to the auxiliary variable.

**Theorem 9.** *Suppose that  $\mathcal{R}$  is of form (24),  $f \in BV_{loc}(\Omega)^n$  and  $u$  is a minimizer of  $\mathcal{E}$  with  $\mathcal{E}(u) < \infty$ . In addition to (H1) and (H2), we assume that*

(H3') *there exists  $v \in \mathcal{D}_2$  such that  $(u, v)$  is a solution to (25) and for any directed inner variation  $\varphi$  there exists a map*

$$\tau \mapsto v_{\varphi, \tau} \in L^1_{loc}(\Omega)^n \quad \text{with } v_{\varphi, 0} = v$$

*defined on a neighbourhood of 0 such that  $\tau \mapsto \tilde{\mathcal{R}}(u_{\tau}^{\varphi}, v_{\varphi, \tau})$  is differentiable at  $\tau = 0$ .*

*If  $\psi$  is  $C^1$  and strictly convex, then  $\mathcal{H}^{m-1}(J_u \setminus J_f) = 0$ . If  $\psi \in C^2$ , then*

$$(u^+ - u^-) \cdot A(u^+ - u^-) \leq (f^+ - f^-) \cdot A(u^+ - u^-) \quad \mathcal{H}^{m-1}\text{-a. e. on } J_u, \quad (26)$$

where

$$A = \int_0^1 D^2\psi(u^- - f^- + s(u^+ - f^+ - u^- + f^+))ds.$$

The proof of this result is identical to the proof of Theorem 1, once the following lemma has been established.

**Lemma 10.** *Let  $u$  be the minimizer of  $\mathcal{E}$  and assume that condition (H3') of Theorem 9 is satisfied. For  $\vartheta \in [0, 1]$ , we denote*

$$u_{\vartheta, \tau}^{\varphi} = \vartheta u_{\tau}^{\varphi} + (1 - \vartheta)u.$$

*Then, for  $\vartheta \in [0, 1]$ ,*

$$\liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, \tau}^{\varphi} - f) - \mathcal{F}(u - f)) + \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, -\tau}^{\varphi} - f) - \mathcal{F}(u - f)) \geq 0.$$

*Proof.* The proof is the same as the proof of Lemma 2, mutatis mutandis. For  $\vartheta \in [0, 1]$ , we denote

$$v_{\varphi, \vartheta, \tau} = \vartheta v_{\varphi, \tau} + (1 - \vartheta)v.$$

By minimality of  $(u, v)$ ,

$$\begin{aligned} 0 &\leq \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\tilde{\mathcal{E}}(u_{\vartheta, \pm\tau}^{\varphi}, v_{\varphi, \vartheta, \pm\tau}) - \tilde{\mathcal{E}}(u, v)) \\ &\leq \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\tilde{\mathcal{R}}(u_{\vartheta, \pm\tau}^{\varphi}, v_{\varphi, \vartheta, \pm\tau}) - \tilde{\mathcal{R}}(u, v) + \mathcal{F}(u_{\vartheta, \pm\tau}^{\varphi} - f) - \mathcal{F}(u - f)). \end{aligned}$$

By our assumption, the function  $\tilde{R}_{\varphi}: ]-\tau_0, \tau_0[ \rightarrow [0, \infty[$  defined by

$$\tilde{R}_{\varphi}(\tau) = \tilde{\mathcal{R}}(u_{\tau}^{\varphi}, v_{\varphi, \tau})$$

for  $\tau_0$  small enough is differentiable at  $\tau = 0$ . Thus, by convexity of  $\tilde{\mathcal{R}}$ ,

$$\frac{1}{\tau} (\tilde{\mathcal{R}}(u_{\vartheta, \pm\tau}^{\varphi}, v_{\varphi, \vartheta, \pm\tau}) - \tilde{\mathcal{R}}(u, v)) \leq \frac{\vartheta}{\tau} (\tilde{\mathcal{R}}(u_{\pm\tau}^{\varphi}, v_{\varphi, \pm\tau}) - \tilde{\mathcal{R}}(u, v)) \rightarrow \pm \vartheta \tilde{R}'_{\varphi}(0) \quad \text{as } \tau \rightarrow 0.$$

Therefore,

$$0 \leq \pm \vartheta \tilde{R}'_{\varphi}(0) + \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, \pm\tau}^{\varphi} - f) - \mathcal{F}(u - f)).$$

We conclude by summing together the two obtained inequalities.  $\square$

## 6.2 Application

Now we will discuss conditions under which regularizers of form (24) satisfy condition  $(H3')$  in the case that

$$\tilde{\mathcal{R}}(w, z) = \mathcal{R}_1(w, z) + \mathcal{R}_2(z), \quad \text{where } \mathcal{R}_1(w, z) = \varrho_1(Dw - z^T \mathcal{L}^m)(\Omega) \quad (27)$$

and  $\mathcal{R}_2: \mathcal{D}_2 \rightarrow [0, \infty]$  is given by one of the following

- $\mathcal{R}_2(z) = TD_{\varrho_2}(z) = \varrho_2(Ez)(\Omega)$ ,  $\mathcal{D}_2 = BD(\Omega)$ ,
- $\mathcal{R}_2(z) = TV_{\varrho_2}(z) = \varrho_2(Dz)(\Omega)$ ,  $\mathcal{D}_2 = BV(\Omega)^m$ ,
- $\mathcal{R}_2(z) = TV_{\varrho_2}(z) = \varrho_2(Dz)(\Omega)$ ,  $\mathcal{D}_2 = \{z \in BV(\Omega)^m : z^T = D\tilde{z}, \tilde{z} \in BV(\Omega)\}$ .

We assume that  $\varrho_1, \varrho_2$  are convex. If  $\varrho_1 = |\cdot|$ ,  $\varrho_2 = |\cdot|$ ,  $\mathcal{R}$  coincides with  $TGV$ ,  $nsTGV$  and  $ICTV$  respectively. However, we are unable to show that  $(H3')$  holds in those cases. Instead, we need to consider partially regularized versions of those functionals. As before, we make the assumption that  $\varrho_1$  satisfies (14), while we make the assumption that  $\varrho_2$  has growth  $p \geq 1$ : there exist  $C_1, C_2$ , with:

$$C_1(|M|^p - 1) \leq \varrho_2(M) \leq C_2(|M|^p + 1) \quad (28)$$

for any  $M \in \mathbb{R}^{m \times m}$ .

**Theorem 11.** *Let  $\tilde{\mathcal{R}}$  be given by (27). Assume that (D1) and (D2) from Theorem 4 hold with  $\varrho = \varrho_1$  and that  $\varrho_2, \varrho_2^{\infty}$  are differentiable. For  $z \in \mathcal{D}_2$ ,  $\varphi \in C_c^{\infty}(\Omega)^m$  and  $\tau$  in a neighborhood of 0 we set*

$$z_{\varphi, \tau}(x) = (I + \tau D\varphi(x))^T z(x + \varphi(x)).$$

*Then for any  $w \in BV(\Omega)$ ,  $z \in \mathcal{D}_2$  the map  $\tau \mapsto \tilde{\mathcal{R}}(w_{\tau}^{\varphi}, z_{\varphi, \tau})$  is differentiable at  $\tau = 0$ .*

*Remark 2.* In case  $p = 1$ , one may assume that  $\varrho_2^\infty$  is differentiable at rank-one matrices only (“rank-one symmetric” matrices [which may have rank 2] for the case of  $TD_{\varrho_2}$ ).

As in the case of Theorem 4, the proof of Theorem 11 relies on change of variables formulae similar to (16).

**Lemma 12.** *Let  $F: \Omega \rightarrow \Omega$  be a diffeomorphism ( $C^2$  up to the boundary) and let  $w \in BV(\Omega)$ ,  $z \in \mathcal{D}_2$ . Then*

$$\mathcal{R}_1(w \circ F, DF^T z \circ F) = \int_{\Omega} |\det D(F^{-1})| d\varrho_1((Dw - z^T \mathcal{L}^m) DF \circ F^{-1}), \quad (29)$$

$$TV_{\varrho_2}(DF^T z \circ F) = \int_{\Omega} |\det D(F^{-1})| d\varrho_2 \left( DF^T \circ F^{-1} Dz DF \circ F^{-1} + D^2 F^T \circ F^{-1} z \mathcal{L}^m \right), \quad (30)$$

$$TD_{\varrho_2}(DF^T z \circ F) = \int_{\Omega} |\det D(F^{-1})| d\varrho_2 \left( DF^T \circ F^{-1} Ez DF \circ F^{-1} + D^2 F^T \circ F^{-1} z \mathcal{L}^m \right). \quad (31)$$

*Proof.* The proof follows along the lines of Lemma 5. In the case of (29), we take as before a sequence  $(w_k) \subset C^\infty(\Omega)$  that converges weakly- $*$  in  $BV(\Omega)$  to  $w$  and show that  $w_k \circ F \xrightarrow{*} w \circ F$  in  $BV(\Omega)$ . We also take  $(z_k) \subset C^\infty(\Omega)^m$  such that  $z_k \rightarrow z$  in  $L^1(\Omega)^m$ ; then  $z_k \circ F \rightarrow z \circ F$  in  $L^1(\Omega)^m$  as well. For any  $h \in C_c(\Omega)^m$ , we calculate

$$\begin{aligned} \int_{\Omega} h \circ F \cdot (D(w_k \circ F) - (DF^T z_k \circ F)^T) &= \int_{\Omega} h \circ F \cdot (Dw_k \circ F - z_k^T \circ F) DF \\ &= \int_{\Omega} h \cdot (Dw_k - z_k^T) DF \circ F^{-1} |\det D(F^{-1})|. \end{aligned}$$

Passing to the limit  $k \rightarrow \infty$ ,

$$\int_{\Omega} h \circ F \cdot d(D(w \circ F) + (DF^T z \circ F)^T \mathcal{L}^m) = \int_{\Omega} |\det D(F^{-1})| h \cdot d(Dw + z^T \mathcal{L}^m) DF \circ F^{-1}.$$

Using (17) as before we deduce (29). Demonstrations of (30) and (31) again follow the same footsteps. In the case of (31), we refer to the proof of [37, Thm. 3.2] which closely follows [18, 1.17], [4] to assert the (weak- $*$ ) density of smooth functions in  $BD(\Omega)$ . In that case, the left multiplication by  $DF^T$  in the change of variable  $DF^T z \circ F$  is crucial to ensure that the symmetrized gradient of the transported function can be estimated in terms of  $Ez$  only, and  $z$  (which is in  $L^{m/(m-1)}$  thanks to Korn–Poincaré’s inequality, see [37, Sec. 1.2]), and does not depend on the skew-symmetric part of  $Dz$ , which is not controlled.  $\square$

*Proof of Theorem 11.* The proof follows along the lines of Theorem 4. In the case of  $ICTV$ -type regularizer we need to note that if  $z = D\tilde{z}^T$  with  $\tilde{z} \in BV(\Omega)$ , then  $DF^T z \circ F = D(\tilde{z} \circ F)^T$ , in particular  $DF^T z \circ F \in \mathcal{D}_2$ . We detail the proof in the  $TD_{\varrho_2}$  case and leave the other cases to the reader. We consider diffeomorphisms of the form  $F_\tau(x) = x + \varphi(x)\nu$  for  $\tau \in \mathbb{R}$  (small),  $\nu$  a unit vector and  $\varphi$  a smooth function with compact support. The term  $\mathcal{R}_1$  will be differentiable as before, so we consider  $\mathcal{R}_2$ , which decomposes as:

$$\begin{aligned} \int_{\Omega} \det D(F_\tau^{-1}(x)) \varrho_2(D^2 F_\tau^T(F_\tau^{-1}(x))z(x) + DF_\tau^T(F_\tau^{-1}(x))e(z(x)DF_\tau(F_\tau^{-1}(x)))) dx \\ + \int_{\Omega} \det D(F_\tau^{-1}(x)) \varrho_2^\infty(DF_\tau(F_\tau^{-1})^T M_z DF_\tau(F_\tau^{-1})) d|E^s z| \quad (32) \end{aligned}$$

where  $M_z$  is the matrix in the polar decomposition of  $Ez$  with respect to  $|Ez|$ .

In case  $p > 1$ , the second integral is not there since  $\varrho_2^\infty \equiv \infty$  and  $E(z)$  is absolutely continuous. On the other hand if  $p = 1$ , that second integral is differentiable at  $\tau = 0$  as soon as  $\varrho_2^\infty$  is differentiable at non-zero rank-one symmetric matrices, since  $M_z$  has such structure  $E^s z$ -a.e. thanks to [12, Thm. 2.3].

Differentiating the first integral is more subtle. Indeed, now, if  $e(z)(x) = 0$ , the term in the absolutely continuous integral does not vanish and is given by  $\det D(F_\tau^{-1})(x) \varrho_2(\tau z(x) \cdot \nu D^2 \varphi(x))$  which is not differentiable if  $\varrho_2$  is not differentiable at 0, for instance in the one-homogeneous case of the standard “TGV”. Assuming that  $\varrho_2$  is  $C^1$ , then, one can write:

$$\begin{aligned} & \int_{\Omega} \det D(F_\tau^{-1}) \varrho_2(D^2 F_\tau^T(F_\tau^{-1})z + DF_\tau(F_\tau^{-1})^T e(z) DF_\tau(F_\tau^{-1})) dx \\ &= \int_{\Omega} \varrho_2(e(z)) dx \\ &+ \int_{\Omega} (\det D(F_\tau^{-1}) - 1) \varrho_2(D^2 F_\tau^T(F_\tau^{-1})z + DF_\tau(F_\tau^{-1})^T e(z) DF_\tau(F_\tau^{-1})) dx \\ &+ \int_0^\tau \int_{\Omega} D\varrho_2\left(e(z) + s(z \cdot \nu D^2 \varphi + 2(e(z)\nu) \odot \nabla \varphi) + s^2(e(z)\nu) \cdot \nu \nabla \varphi \otimes \nabla \varphi\right) \\ &\quad \cdot \left(z \cdot \nu D^2 \varphi + 2(e(z)\nu) \odot \nabla \varphi + 2s(e(z)\nu) \cdot \nu \nabla \varphi \otimes \nabla \varphi\right) dx ds, \end{aligned}$$

where the notation  $a \odot b$  stands for the symmetric tensor product  $(a \otimes b + b \otimes a)/2$ . For any two matrices  $A, B$ ,

$$\varrho_2(A \pm B) - \varrho_2(A) \geq \pm D\varrho_2(A) \cdot B$$

and one deduces from the growth assumption (28) that there is  $C > 0$  such that

$$|D\varrho_2(A) \cdot B| \leq C(|A|^p + |B|^p + 1).$$

This allows to bound the integrand in the last formula by  $C'(1 + |e(z)|^p + |z|^p) \in L^1(\Omega)$  (again, thanks to Korn or Poincaré–Korn’s inequality) for some constant  $C' > 0$ , and apply Lebesgue’s dominated convergence to deduce that (32) is differentiable at  $\tau = 0$ .  $\square$

*Remark 3.* If  $\varrho_2$  is 1-homogeneous, we do now know whether the result holds. It is however likely that the condition  $(H\mathfrak{S}')$  is not general enough to lead to a conclusion, and that one might need a more complicated decomposition of the functions, as suggested in [39]. On the other hand the result in the cases  $p > 1$  is already proved in that reference.

## 7 Data of unbounded variation

In this section, we discuss the case where  $f \notin BV(\Omega)^m$  and we only address the simplest case where the data term is strongly convex and with Lipschitz gradient, that is, verifies (8). We introduce a weaker description of a “jump set” (which for  $BV$  functions coincides with the standard jump set up to a negligible set), for which we are still able to deduce jump inclusion. For  $f \in L^2_{loc}(\Omega)^n$ ,  $x_0 \in \Omega$ ,  $\nu_0 \in \mathbb{S}^{m-1}$  we define

$$j_{f,\nu_0}(x_0)^2 := \limsup_{\tau \rightarrow 0^+} \int_{Q_\tau^-(x_0, \nu_0)} |f(x + \tau \nu_0) - f(x)|^2, \quad j_f(x_0) := \sup_{\nu_0 \in \mathbb{S}^{m-1}} j_{f,\nu_0}(x_0).$$

(See (2) for the notation  $Q_r^\pm(x, \nu)$ .) We denote by  $\tilde{J}_f$  the set of  $x_0 \in \Omega$  such that  $j_f(x_0) > 0$ .

**Proposition 13.** For  $f \in L_{loc}^2(\Omega)^n$ ,  $x_0 \in \Omega$ ,  $\nu_0 \in \mathbb{S}^{m-1}$  we have

$$j_f(x_0, \nu_0)^2 \leq 4 \limsup_{\tau \rightarrow 0^+} \int_{Q_\tau(x_0, \nu_0)} \left| f - \int_{Q_\tau(x_0, \nu_0)} f \right|^2.$$

In particular, if  $x_0$  is a (2-)Lebesgue point of  $f$ , then  $j_f(x_0) = 0$ . If  $f \in L_{loc}^\infty(\Omega)^n$ , then

$$J_f \subset \tilde{J}_f \subset S_f.$$

Moreover, for  $x_0 \in J_f$ ,  $j_f(x_0) = j_{f, \nu_f(x_0)}(x_0) = |f^+ - f^-|(x_0)$ .

See Section 2.2 or [3, Sec. 3] for the definition of  $J_f$ ,  $S_f$ .

*Proof.* First of all, we indeed have

$$\begin{aligned} j_{f, \nu_0}(x_0)^2 &\leq 2 \limsup_{\tau \rightarrow 0^+} \int_{Q_\tau^-(x_0, \nu_0)} \left| f(x + \tau \nu_0) - \int_{Q_\tau(x_0, \nu_0)} f(x) \right|^2 + \left| f(x) - \int_{Q_\tau(x_0, \nu_0)} f(x) \right|^2 \\ &= 2 \limsup_{\tau \rightarrow 0^+} \int_{Q_\tau^+(x_0, \nu_0)} \left| f(x) - \int_{Q_\tau(x_0, \nu_0)} f(x) \right|^2 + \int_{Q_\tau^-(x_0, \nu_0)} \left| f(x) - \int_{Q_\tau(x_0, \nu_0)} f(x) \right|^2 \\ &= 4 \limsup_{\tau \rightarrow 0^+} \int_{Q_\tau(x_0, \nu_0)} \left| f - \int_{Q_\tau(x_0, \nu_0)} f \right|^2. \end{aligned}$$

It is known that Lebesgue points calculated with respect to different *regular families* of sets are the same. In particular in our case, observing that  $B_{\sqrt{2}\tau}(x_0) \supseteq Q_\tau(x_0, \nu_0)$ , we have for every  $\nu_0 \in \mathbb{S}^{m-1}$

$$\begin{aligned} \int_{Q_\tau(x_0, \nu_0)} \left| f - \int_{Q_\tau(x_0, \nu_0)} f \right|^2 &\leq 2 \int_{Q_\tau(x_0, \nu_0)} \left| f - \int_{B_{\sqrt{2}\tau}(x_0)} f \right|^2 + 2 \left| \int_{Q_\tau(x_0, \nu_0)} f - \int_{B_{\sqrt{2}\tau}(x_0)} f \right|^2 \\ &\leq 4 \int_{Q_\tau(x_0, \nu_0)} \left| f - \int_{B_{\sqrt{2}\tau}(x_0)} f \right|^2 \leq 4 \frac{\mathcal{L}^m(B_{\sqrt{2}\tau}(x_0))}{\mathcal{L}^m(Q_\tau(x_0, \nu_0))} \int_{B_{\sqrt{2}\tau}(x_0)} \left| f - \int_{B_{\sqrt{2}\tau}(x_0)} f \right|^2 \end{aligned}$$

Thus, if  $x_0$  is a (2-)Lebesgue point of  $f$ , then  $j_{f, \nu_0}(x_0) = 0$  for every  $\nu_0 \in \mathbb{S}^{m-1}$ , whence  $j_f(x_0) = 0$ .

Now suppose that  $f \in L_{loc}^\infty(\Omega)^n$ . If  $x_0 \in \Omega \setminus S_f$ , then

$$\int_{B_\tau(x_0)} \left| f - \int_{B_\tau(x_0)} f \right|^2 \leq 2 \|f\|_{L^\infty(B_\tau(x_0))} \int_{B_\tau(x_0)} \left| f - \int_{B_\tau(x_0)} f \right| \rightarrow 0 \quad \text{as } \tau \rightarrow 0^+,$$

so  $j_f(x_0) = 0$ , i. e.  $x_0 \in \Omega \setminus \tilde{J}_f$ . On the other hand, if  $x_0 \in J_f$  and  $\nu_0$  is the direction of jump of  $f$  at  $x_0$ , then by the triangle inequality in  $L^2(Q_\tau^-(x_0, \nu_0))^n$ ,

$$\begin{aligned} &\sqrt{\int_{Q_\tau^-(x_0, \nu_0)} |f(x + \tau \nu_0) - f(x)|^2} \\ &\geq -\sqrt{\int_{Q_\tau^+(x_0, \nu_0)} |f - f^+(x_0)|^2} + |f^+(x_0) - f^-(x_0)| - \sqrt{\int_{Q_\tau^-(x_0, \nu_0)} |f - f^-(x_0)|^2}. \end{aligned}$$

Since

$$\int_{Q_\tau^\pm(x_0, \nu_0)} |f - f^\pm(x_0)|^2 \leq 2\|f\|_{L^\infty(Q_\tau^\pm(x_0, \nu_0))}^n \int_{Q_\tau^\pm(x_0, \nu_0)} |f - f^\pm(x_0)| \rightarrow 0 \quad \text{as } \tau \rightarrow 0^+,$$

we obtain

$$j_f(x_0) \geq j_{f, \nu_0}(x_0) \geq |f^+(x_0) - f^-(x_0)|,$$

in particular  $x_0 \in \tilde{J}_f$ . It remains to prove the opposite inequality. Let  $\nu \in \mathbb{S}^{m-1}$  and let  $q = q(\nu_0, \nu) \geq 1$  be the smallest number such that  $Q_\tau(x_0, \nu) \subset Q_{q\tau}(x_0, \nu_0)$ . We stress that  $q$  does not depend on  $\tau$ . Assume without loss of generality that  $\nu_0 \cdot \nu \geq 0$ . Then  $Q_\tau^\pm(x_0, \nu)$  can be divided into six parts (see Figure 1):

$$A_\tau^{0,\pm}(x_0, \nu) = \{x \in Q_\tau^\pm(x_0, \nu) : x \in Q_{q\tau}^\pm(x_0, \nu_0), x \mp \tau\nu \in Q_{q\tau}^\mp(x_0, \nu_0)\},$$

$$A_\tau^{+,\pm}(x_0, \nu) = \{x \in Q_\tau^\pm(x_0, \nu) : x \in Q_{q\tau}^\pm(x_0, \nu_0), x \mp \tau\nu \in Q_{q\tau}^\pm(x_0, \nu_0)\},$$

$$A_\tau^{-,\pm}(x_0, \nu) = \{x \in Q_\tau^\pm(x_0, \nu) : x \in Q_{q\tau}^\mp(x_0, \nu_0), x \mp \tau\nu \in Q_{q\tau}^\mp(x_0, \nu_0)\}.$$

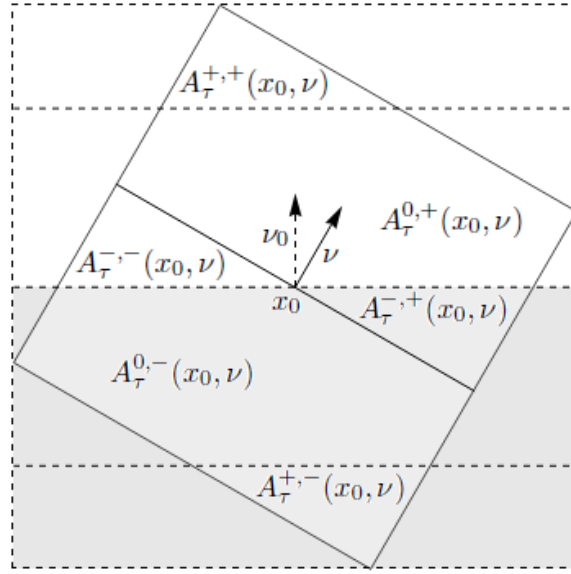


Figure 1: Sets  $A_\tau^{0,\pm}(x_0, \nu)$ ,  $A_\tau^{+,\pm}(x_0, \nu)$ ,  $A_\tau^{-,\pm}(x_0, \nu)$ .

By definition,  $A_\tau^{\mp,+}(x_0, \nu) \cup A_\tau^{\pm,-}(x_0, \nu) \subset Q_{q\tau}^\mp(x_0, \nu_0)$ , so

$$\begin{aligned} \int_{A_\tau^{\pm,-}(x_0, \nu)} |f(x + \tau\nu_0) - f(x)|^2 &= \int_{A_\tau^{\pm,-}(x_0, \nu)} |f(x + \tau\nu_0) - f^\mp(x_0) + f^\mp(x_0) - f(x)|^2 \\ &\leq 2 \int_{A_\tau^{\mp,+}(x_0, \nu)} |f - f^\mp(x_0)|^2 + 2 \int_{A_\tau^{\pm,-}(x_0, \nu)} |f - f^\mp(x_0)|^2 = 2 \int_{A_\tau^{\mp,+}(x_0, \nu) \cup A_\tau^{\pm,-}(x_0, \nu)} |f - f^\mp(x_0)|^2 \\ &\leq 4\|f\|_{L^\infty(Q_{q\tau}^\mp(x_0, \nu_0))}^n \int_{Q_{q\tau}^\mp(x_0, \nu_0)} |f - f^\mp(x_0)| \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{\mathcal{L}^m(Q_\tau^-(x_0, \nu))} \int_{A_\tau^{\pm, -}(x_0, \nu)} |f(x + \tau\nu_0) - f(x)|^2 \\ & \leq 4\|f\|_{L^\infty(Q_{q\tau}^\mp(x_0, \nu_0))}^n q^m \int_{Q_{q\tau}^\mp(x_0, \nu_0)} |f - f^\mp(x_0)| \rightarrow 0 \text{ as } \tau \rightarrow 0^+. \end{aligned}$$

Therefore,

$$\begin{aligned} \limsup_{\tau \rightarrow 0^+} \int_{Q_\tau^-(x_0, \nu_0)} |f(x + \tau\nu_0) - f(x)|^2 &= \limsup_{\tau \rightarrow 0^+} \frac{1}{\mathcal{L}^m(Q_\tau^-(x_0, \nu))} \int_{A_\tau^{0, -}(x_0, \nu)} |f(x + \tau\nu_0) - f(x)|^2 \\ &\leq \limsup_{\tau \rightarrow 0^+} \int_{A_\tau^{0, -}(x_0, \nu)} |f(x + \tau\nu_0) - f(x)|^2. \end{aligned}$$

Then, by the triangle inequality in  $L^2(A_\tau^{0, -}(x_0, \nu))^n$ ,

$$\begin{aligned} & \sqrt{\int_{A_\tau^{0, -}(x_0, \nu)} |f(x + \tau\nu_0) - f(x)|^2} \\ & \leq \sqrt{\int_{A_\tau^{0, +}(x_0, \nu)} |f - f^+(x_0)|^2} + |f^+(x_0) - f^-(x_0)| + \sqrt{\int_{A_\tau^{0, -}(x_0, \nu)} |f - f^-(x_0)|^2}. \end{aligned}$$

We estimate

$$\int_{A_\tau^{0, \pm}(x_0, \nu)} |f - f^\pm(x_0)|^2 \leq 2\|f\|_{L^\infty(A_\tau^{0, \pm}(x_0, \nu))}^n \frac{\mathcal{L}^m(Q_{q\tau}^\pm(x_0, \nu_0))}{\mathcal{L}^m(A_\tau^{0, \pm}(x_0, \nu))} \int_{Q_{q\tau}^\pm(x_0, \nu_0)} |f - f^\pm(x_0)|.$$

Since the quotient  $\mathcal{L}^m(Q_{q\tau}^\pm(x_0, \nu_0))/\mathcal{L}^m(A_\tau^{0, \pm}(x_0, \nu))$  is independent of  $\tau$ , the r. h. s. converges to 0 as  $\tau \rightarrow 0^+$ , whence

$$j_{f, \nu}(x_0) \leq |f^+(x_0) - f^-(x_0)|.$$

As  $\nu \in \mathbb{S}^{m-1}$  is arbitrary, we conclude.  $\square$

As a consequence of Proposition 13 and the Federer–Vol’pert Theorem [3, Theorem 3.78], if  $f \in L_{loc}^\infty(\Omega)^n \cap BV_{loc}(\Omega)^n$ , the three sets  $S_f$ ,  $\tilde{J}_f$  and  $J_f$  coincide up to  $\mathcal{H}^{m-1}$ -negligible sets.

**Theorem 14.** *Let  $f \in L^\infty(\Omega)^n$ , suppose that  $\mathcal{E}$  admits a minimizer  $u \in L^\infty(\Omega)^n \cap BV(\Omega)^n$ ,  $\psi$  is  $C^2$  and (8) holds on  $\{z \in \mathbb{R}^n : |z| \leq \|u\|_{L^\infty(\Omega)^n} + \|f\|_{L^\infty(\Omega)^n}\}$ . Assume (H3) or that  $\mathcal{R}$  is of form (24) and (H3’) holds. Then  $J_u \subset \tilde{J}_f$  up to a  $\mathcal{H}^{m-1}$ -negligible set and*

$$|u^+ - u^-|(x_0) \leq \sqrt{\Lambda/\lambda} j_f(x_0) \quad \text{for } \mathcal{H}^{m-1}\text{-a. e. } x_0 \in J_u.$$

*Proof.* By Lemma 2 (or Lemma 10 in the inf-convolution setting) we have for any (directed) inner variation  $\varphi$  and  $\vartheta \in [0, 1]$

$$0 \leq \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, \tau}^\varphi - f) - \mathcal{F}(u - f)) + \frac{1}{\tau} (\mathcal{F}(u_{\vartheta, -\tau}^\varphi - f) - \mathcal{F}(u - f)). \quad (33)$$

We take  $\Gamma$ ,  $x_0$ ,  $\nu_0 = \nu_u(x_0)$  and  $r_0$  as in the beginning of the proof of Theorem 1, except now we cannot assume that the traces  $f^\pm$  exist on both sides of  $\Gamma$ . Instead we assume that  $x_0$  is a Lebesgue point of  $j_f$  with respect to  $\mathcal{H}^{m-1} \llcorner \Gamma$ , in particular

$$\int_{Q_r \cap \Gamma} j_f(x)^2 d\mathcal{H}^{m-1}(x) \xrightarrow{r \rightarrow 0} j_f(x_0)^2. \quad (34)$$



As in the proof of Lemma 3, by an isometric change of coordinates, we assume that  $\nu_0 = e_m$ ,  $x_0 = 0$  and denote  $x = (x', x_m)$ ,  $Q_r(x_0, \nu_0) = Q_r$ ,  $B_r^{m-1}(x_0, \nu_0) = B_r^{m-1}$ ,

$$\Gamma = \left\{ \gamma(x') : x' \in B_r^{m-1} \right\},$$

$\gamma(x') = (x', \tilde{\gamma}(x'))$ . We recall that we assume  $\tilde{\gamma}$  is  $C^1$ . For  $s \leq r$ , we let  $L_s = \max_{B_s^{m-1}} |\nabla' \tilde{\gamma}|$  the Lipschitz constant of  $\tilde{\gamma}$  on  $B_s^{m-1}$ , which is such that  $\lim_{s \rightarrow 0} L_s = 0$ .

Given  $0 < s < r$ , we take  $\varphi \in C_c^\infty(\Omega)^n$  such that the support of  $\varphi$  is contained in  $Q_r$  and  $\varphi = \nu_0 \tilde{\varphi}$ , where  $0 \leq \tilde{\varphi} \leq 1$  and  $\tilde{\varphi} = 1$  on  $\overline{B_s^{m-1}} \times [-r/2, r/2]$ . We denote

$$S_{s,\tau}^- = \{(x', x_m) : x' \in \overline{B_s^{m-1}}, \tilde{\gamma}(x') - \tau \leq x_m \leq \tilde{\gamma}(x')\},$$

$$S_{s,\tau}^+ = \{(x', x_m) : x' \in \overline{B_s^{m-1}}, \tilde{\gamma}(x') \leq x_m \leq \tilde{\gamma}(x') + \tau\},$$

$$u_\tau(x) = u(x', x_m + \tau), \quad f_\tau(x) = f(x', x_m + \tau),$$

$$u_{\vartheta,\tau} = \vartheta u_\tau + (1 - \vartheta)u, \quad f_{\vartheta,\tau} = \vartheta f_\tau + (1 - \vartheta)f.$$

We note that  $u_\tau = u_\tau^\varphi$  and  $u_{\vartheta,\tau} = u_{\vartheta,\tau}^\varphi$  on  $S_{s,\tau}^-$  as soon as  $\tau \leq r/2$ . We decompose

$$\mathcal{F}(u_{\vartheta,\pm\tau}^\varphi - f) - \mathcal{F}(u - f) = \int_{S_{s,\tau}^\mp} \psi(u_{\vartheta,\pm\tau}^\varphi - f) - \psi(u - f) + \int_{Q_r \setminus S_{s,\tau}^\mp} \psi(u_{\vartheta,\pm\tau}^\varphi - f) - \psi(u - f).$$

Reasoning as in the proof of Lemma 3 (where we use that  $\psi$  is Lipschitz in  $\{z \in \mathbb{R}^n : |z| \leq \|u\|_{L^\infty(\Omega)^n} + \|f\|_{L^\infty(\Omega)^n}\}$ ), we get

$$\frac{1}{\tau} \int_{Q_r \setminus S_{s,\tau}^\mp} \psi(u_{\vartheta,\pm\tau}^\varphi - f) - \psi(u - f) \leq \frac{C}{\tau} \int_{Q_r \setminus S_{s,\tau}^\mp} |u_{\vartheta,\pm\tau}^\varphi - u| \leq C\vartheta|\nu_0 \cdot Du|(Q_r \setminus (Q_s \cap \Gamma)).$$

On the other hand, using the convexity of  $\psi$ ,

$$\begin{aligned} & \int_{S_{s,\tau}^-} \psi(u_{\vartheta,\tau}^\varphi - f) - \psi(u - f) + \int_{S_{s,\tau}^+} \psi(u_{\vartheta,-\tau}^\varphi - f) - \psi(u - f) \\ &= \int_{S_{s,\tau}^-} \psi(u_{\vartheta,\tau} - f) - \psi(u - f) + \psi(u_{(1-\vartheta),\tau} - f_\tau) - \psi(u_\tau - f_\tau) \\ &\leq \int_{S_{s,\tau}^-} \vartheta D\psi(u_{\vartheta,\tau} - f) \cdot (u_\tau - u) + \vartheta D\psi(u_{(1-\vartheta),\tau} - f_\tau) \cdot (u - u_\tau) \\ &= \vartheta \int_{S_{s,\tau}^-} (D\psi(u_{\vartheta,\tau} - f) - D\psi(u_{(1-\vartheta),\tau} - f_\tau)) \cdot (u_\tau - u) \\ &= \vartheta \int_{S_{s,\tau}^-} (u_{\vartheta,\tau} - u_{(1-\vartheta),\tau} - f + f_\tau) \cdot A_{\vartheta,\tau} \cdot (u_\tau - u), \end{aligned}$$

where the symmetric positive definite matrix  $A_{\vartheta,\tau}$  is given by:

$$A_{\vartheta,\tau} = \int_0^1 D^2\psi(u_{(1-\vartheta),\tau} - f_\tau + s(u_{\vartheta,\tau} - u_{(1-\vartheta),\tau} - f + f_\tau)) ds.$$

Using  $2\xi \cdot A_{\vartheta,\tau} \cdot \eta \leq \xi \cdot A_{\vartheta,\tau} \cdot \xi + \eta \cdot A_{\vartheta,\tau} \cdot \eta$  for any  $\xi, \eta$  and (8) it follows that

$$\begin{aligned} & (u_{\vartheta,\tau} - u_{(1-\vartheta),\tau} - f + f_\tau) \cdot A_{\vartheta,\tau} \cdot (u_\tau - u) \\ &= -(1-2\vartheta)(u_\tau - u) \cdot A_{\vartheta,\tau} \cdot (u_\tau - u) + (f_\tau - f) \cdot A_{\vartheta,\tau} \cdot (u_\tau - u) \\ &\leq -\left(\frac{1}{2} - 2\vartheta\right)(u_\tau - u) \cdot A_{\vartheta,\tau} \cdot (u_\tau - u) + \frac{1}{2}(f_\tau - f) \cdot A_{\vartheta,\tau} \cdot (f_\tau - f) \\ &\leq -\left(\frac{1}{2} - 2\vartheta\right)\lambda|u_\tau - u|^2 + \frac{1}{2}\Lambda|f_\tau - f|^2 \end{aligned}$$

for  $\vartheta \in [0, \frac{1}{4}]$ . Using [3, Theorem 3.108], recalling (33) and combining the estimates above,

$$\begin{aligned} \vartheta(1-4\vartheta)\lambda \int_{Q_s \cap \Gamma} |u^+ - u^-|^2 d\gamma_{\#} \mathcal{L}^{m-1} &= \vartheta(1-4\vartheta)\lambda \lim_{\tau \rightarrow 0^+} \frac{1}{\tau} \int_{S_{s,\tau}^-} |u_\tau - u|^2 \\ &\leq \vartheta\Lambda \liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} \int_{S_{s,\tau}^-} |f_\tau - f|^2 + 2C\vartheta|\nu_0 \cdot Du|(Q_r \setminus (Q_s \cap \Gamma)). \end{aligned} \quad (35)$$

We recall that here,  $u^\pm$  coincide with the traces of  $u$  on both side of  $\Gamma$ . We estimate the pushforward measure  $\gamma_{\#} \mathcal{L}^{m-1}$  by

$$\gamma_{\#} \mathcal{L}^{m-1} \llcorner (Q_s \cap \Gamma) = \frac{\mathcal{H}^{m-1} \llcorner (Q_s \cap \Gamma)}{\sqrt{1 + |\nabla' \gamma(x')|^2}} \geq \frac{1}{\sqrt{1 + L_s^2}} \mathcal{H}^{m-1} \llcorner (Q_s \cap \Gamma).$$

Now, for  $x' \in B_s^{m-1}$ ,  $\tau < r - s$ , let us denote

$$\tilde{Q}_\tau^-(x') = \{(y', y_m) \in Q_r : |y' - x'| < \tau, \tilde{\gamma}(y') - \tau < y_m < \tilde{\gamma}(y')\}.$$

We observe that  $\mathcal{L}^m(\tilde{Q}_\tau^-(x')) = \tau \mathcal{L}^{m-1}(B_\tau^{m-1})$ . Then

$$\frac{1}{\tau} \int_{S_{s,\tau}^-} |f_\tau - f|^2 \leq \int_{B_s^{m-1}} \int_{\tilde{Q}_\tau^-(x')} |f_\tau - f|^2.$$

Note that  $\int_{\tilde{Q}_\tau^-(x')} |f_\tau - f|^2$  is uniformly bounded by  $4\|f\|_{L^\infty(Q_r)}^2$ . Thus, by Fatou's Lemma

$$\liminf_{\tau \rightarrow 0^+} \frac{1}{\tau} \int_{S_{s,\tau}^-} |f_\tau - f|^2 \leq \int_{B_s^{m-1}} \limsup_{\tau \rightarrow 0^+} \int_{\tilde{Q}_\tau^-(x')} |f_\tau - f|^2.$$

Dividing (35) by  $\vartheta$  and passing to the limits  $\vartheta \rightarrow 0^+$ ,  $s \rightarrow r^-$

$$\lambda \int_{Q_r \cap \Gamma} |u^+ - u^-|^2 d\gamma_{\#} \mathcal{L}^{m-1} \leq \Lambda \int_{B_r^{m-1}} \limsup_{\tau \rightarrow 0^+} \int_{\tilde{Q}_\tau^-(x')} |f_\tau - f|^2 + 2C|\nu_0 \cdot Du|(Q_r \setminus \Gamma). \quad (36)$$

We estimate

$$\mathcal{L}^m(\tilde{Q}_\tau^-(x') \setminus Q_\tau^-(x', \gamma(x'), \nu_0)) \leq \mathcal{L}^{m-1}(B_\tau^{m-1}) \tau \max_{B_r^{m-1}} |\nabla' \gamma| = L_r \mathcal{L}^m(\tilde{Q}_\tau^-(x')) \tau.$$

Therefore,

$$\int_{\tilde{Q}_\tau^-(x')} |f_\tau - f|^2 \leq \int_{Q_\tau^-(x', \gamma(x'), \nu_0)} |f_\tau - f|^2 + 4L_r \|f\|_{L^\infty(Q_r)}^2$$

and

$$\limsup_{\tau \rightarrow 0^+} \int_{\tilde{Q}_\tau^-(x')} |f_\tau - f|^2 \leq j_f(x', \gamma(x'))^2 + 4L_r \|f\|_{L^\infty(Q_r)}^2.$$

Hence, we can estimate

$$\begin{aligned}
& \int_{B_r^{m-1}} \limsup_{\tau \rightarrow 0^+} \int_{\tilde{Q}_\tau^-(x')} |f_\tau - f|^2 \\
& \leq \int_{B_r^{m-1}} j_f(x', \gamma(x'))^2 d\mathcal{L}^{m-1}(x') + 4\mathcal{L}^{m-1}(B_r^{m-1})L_r \|f\|_{L^\infty(Q_r)^n}^2 \\
& = \int_{Q_r \cap \Gamma} \frac{j_f(x)^2}{\sqrt{1 + |\nabla' \tilde{\gamma}(x')|^2}} d\mathcal{H}^{m-1}(x) + 4\mathcal{L}^{m-1}(B_r^{m-1})L_r \|f\|_{L^\infty(Q_r)^n}^2 \\
& \leq \int_{Q_r \cap \Gamma} j_f^2 d\mathcal{H}^{m-1} + 4\mathcal{L}^{m-1}(B_r^{m-1})L_r \|f\|_{L^\infty(Q_r)^n}^2
\end{aligned}$$

Recalling (36), we deduce

$$\begin{aligned}
\lambda \int_{Q_r \cap \Gamma} |u^+ - u^-|^2 d\gamma_{\#} \mathcal{L}^{m-1} & \leq \Lambda \int_{Q_r \cap \Gamma} j_f^2 d\mathcal{H}^{m-1} \\
& \quad + 4\mathcal{L}^{m-1}(B_r^{m-1})L_r \|f\|_{L^\infty(Q_r)^n}^2 + 2C|\nu_0 \cdot Du|(Q_r \setminus \Gamma). \quad (37)
\end{aligned}$$

Finally, we divide both sides of (37) by  $\mathcal{L}^{m-1}(B_r^{m-1}) = \gamma_{\#} \mathcal{L}^{m-1}(Q_r \cap \Gamma)$  and keeping in mind that  $\mathcal{H}^{m-1}(Q_r \cap \Gamma)/\mathcal{L}^{m-1}(B_r^{m-1}) \rightarrow 1$  and  $L_r \rightarrow 0$  as  $r \rightarrow 0^+$ , we pass to the limit obtaining the asserted inequality owing to (34).  $\square$

## 8 Experiment



Figure 2: A noisy image and the denoised versions with, respectively, the Frobenius ROF, Nuclear ROF, Spectral ROF problems.

We solved here the “ROF” problem (1) for a data term given by a noisy color image, and the Frobenius, Nuclear and Spectral total variations. Figure 2 shows the results, which look almost identical. The close-up in Figure 3 seems to show that the edges are better recovered with the Nuclear total variation, and quite jagged in the case of the Spectral total variation, for which the jump inclusion might not hold. Of course, this is a discrete experiment at a fixed scale and therefore a relatively poor illustration of our main results. Observe that in these experiments, one cannot expect that the original datum (left image) represents a function



Figure 3: Detail of Figure 2

$f \in BV(\Omega, [0, 1]^3)$ . However, being obtained by adding a small amplitude noise to a bounded variation function, we may expect that the set  $\tilde{J}_f$  of Section 7 corresponds to the set of (large enough) edges in the original image.

## 9 Conclusion and comments

We have introduced an approach for the study of the jump set of minimizers of “Rudin-Osher-Fatemi” type problems which is more versatile than the original approach in [8], and easier to handle than [38] (even if equivalent in spirit, and much inspired by it). We recover many cases (and more) from the previous works [38, 39], including jump inclusion in the “TGV” case (up to some smoothing). The full nonsmooth case remains open. Also, our approach does not seem to allow to derive further regularity, such as the continuity results of [31, 9]. It is also unclear what exactly happens for non-differentiable norms, since at jumps the gradients have rank one, and locally many nonsmooth norms remain differentiable—yet experimental observations seem to show a much worse control on the oscillations parallel to the jumps in such cases, as in Figure 3 (right).

Our results can be rephrased in terms of the measure  $D^j u$ . In particular, in the case of  $\psi$  strongly convex with Lipschitz-gradient (8), Theorems 1 and 9 imply estimate  $|D^j u| \leq \sqrt{\lambda/\Lambda} |D^j f|$ . Similar bounds have been recently obtained for the whole singular part  $D^s u$  in the 1D vectorial case ( $m = 1, n > 1$ ) in [20]. To our knowledge, it remains an open question whether such estimates hold for the Cantor part  $D^c u$  in  $m > 1$ , even in the case of scalar  $TV$ , although in  $n = 1$  it is known that  $D^s f = 0 \implies D^s u = 0$  for general regularizers of form (15) if  $\Omega$  is convex [27], and that regularity away from the jump set is transferred to the solution (for  $\mathcal{R}$  nice enough) [9, 31].

Finally, a natural question is whether the results shown in this work also hold for the gradient flow of the total variation or similar functionals. In [8, 10], this is deduced from Crandall–Liggett’s theorem in  $L^\infty(\Omega)$ , which can be applied because minimizing the ROF problem (1) is contractive in the sup norm. Yet, this is unknown (and possibly not true) in the vectorial case and no easy conclusion may be drawn. In relation to this, we mention a recent paper [25], where an interesting continuity property of the map  $w \mapsto D^j w$  is obtained. However, its applicability in our context remains a matter of further investigation.

## Acknowledgements

The authors are indebted to Rémy Rodiac for inspiring discussions about differentiability along inner variations, and to Irène Waldspurger for sorting out the properties of Schatten norms. ML was supported by grant 2020/36/C/ST1/00492 of the National Science Centre, Poland.

## References

- [1] Giovanni Alberti. Rank one property for derivatives of functions with bounded variation. *Proc. Roy. Soc. Edinburgh Sect. A* 123 (1993) pp. 239–274. doi:10.1017/S030821050002566X.
- [2] Luigi Ambrosio, Gianni Dal Maso. A general chain rule for distributional derivatives. *Proc. Amer. Math. Soc.* 108 (1990) pp. 691–702. doi:10.2307/2047789.
- [3] Luigi Ambrosio, Nicola Fusco, Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York (2000).
- [4] Gabriele Anzellotti, Mariano Giaquinta. Funzioni  $BV$  e tracce. *Rend. Sem. Mat. Univ. Padova* 60 (1978) pp. 1–21.
- [5] Rajendra Bhatia. *Perturbation bounds for matrix eigenvalues*, vol. 53 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2007). doi:10.1137/1.9780898719079. Reprint of the 1987 original.
- [6] Kristian Bredies, Karl Kunisch, Thomas Pock. Total generalized variation. *SIAM J. Imaging Sci.* 3 (2010) pp. 492–526. doi:10.1137/090769521.
- [7] Antoni Buades, Bartomeu Coll, Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* 4 (2005) pp. 490–530. doi:10.1137/040616024.
- [8] Vicent Caselles, Antonin Chambolle, Matteo Novaga. The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model. Simul.* 6 (2007) pp. 879–894. doi:10.1137/070683003.
- [9] Vicent Caselles, Antonin Chambolle, Matteo Novaga. Regularity for solutions of the total variation denoising problem. *Rev. Mat. Iberoam.* 27 (2011) pp. 233–252. doi:10.4171/RMI/634.
- [10] Vicent Caselles, Khalid Jalalzai, Matteo Novaga. On the jump set of solutions of the total variation flow. *Rend. Semin. Mat. Univ. Padova* 130 (2013) pp. 155–168. doi:10.4171/RSMUP/130-5.
- [11] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16 (2007) pp. 2080–2095. doi:10.1109/TIP.2007.901238.

- [12] Guido De Philippis, Filip Rindler. On the structure of measures constrained by linear PDEs. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. III. Invited lectures*. World Sci. Publ., Hackensack, NJ (2018) pp. 2215–2239.
- [13] Giacomo Del Nin. Rectifiability of the jump set of locally integrable functions. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 22 (2021) pp. 1233–1240.
- [14] Françoise Demengel, Roger Temam. Convex functions of a measure and applications. *Indiana Univ. Math. J.* 33 (1984) pp. 673–709. doi:10.1512/iumj.1984.33.33036.
- [15] Joan Duran, Michael Moeller, Catalina Sbert, Daniel Cremers. Collaborative total variation: a general framework for vectorial TV models. *SIAM J. Imaging Sci.* 9 (2016) pp. 116–151. doi:10.1137/15M102873X.
- [16] Michael Elad, Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* 15 (2006) pp. 3736–3745. doi:10.1109/TIP.2006.881969.
- [17] Lawrence C. Evans, Ronald F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL (1992).
- [18] Enrico Giusti. *Minimal surfaces and functions of bounded variation*, vol. 80 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel (1984). doi:10.1007/978-1-4684-9486-0.
- [19] Bastian Goldluecke, Evgeny Strekalovskiy, Daniel Cremers. The natural vectorial total variation which arises from geometric measure theory. *SIAM J. Imaging Sci.* 5 (2012) pp. 537–564. doi:10.1137/110823766.
- [20] Zofia Grochulska, Michał Łasica. Local estimates for vectorial Rudin-Osher-Fatemi type problems in one dimension (2022). [arXiv:2209.05415](https://arxiv.org/abs/2209.05415).
- [21] Viren Jain, Sebastian Seung. Natural image denoising with convolutional networks. In D. Koller, D. Schuurmans, Y. Bengio, L. Bottou, eds., *Advances in Neural Information Processing Systems*, vol. 21. Curran Associates, Inc. (2008) .
- [22] Khalid Jalalzai. *Regularization of inverse problems in image processing*. Theses, Ecole Polytechnique X (2012). Rédaction : fin 2011. Soutenance : mars 2012.
- [23] Khalid Jalalzai. Discontinuities of the minimizers of the weighted or anisotropic total variation for image reconstruction (2014). [arXiv:1402.0026](https://arxiv.org/abs/1402.0026).
- [24] Khalid Jalalzai, Antonin Chambolle. Properties of minimizers of the total variation and of the solutions of the total variation flow. In *2014 IEEE International Conference on Image Processing (ICIP)* (2014) pp. 4832–4836. doi:10.1109/ICIP.2014.7025979.
- [25] Krystian Kazaniecki, Anton Tselishchev, Michał Wojciechowski. Schur property for jump parts of gradient measures (2023). [arXiv:2307.08396](https://arxiv.org/abs/2307.08396).
- [26] Michał Łasica, Salvador Moll, Piotr B. Mucha. Total variation denoising in  $l^1$  anisotropy. *SIAM J. Imaging Sci.* 10 (2017) pp. 1691–1723. doi:10.1137/16M1103610.

- [27] Michał Łasica, Piotr Rybka. Existence of  $W^{1,1}$  solutions to a class of variational problems with linear growth on convex domains. *Indiana Univ. Math. J.* 70 (2021) pp. 2427–2450. doi:10.1512/iumj.2021.70.9479.
- [28] Marc Lebrun, Miguel Colom, Antoni Buades, Jean-Michel Morel. Secrets of image denoising cuisine. *Acta Numer.* 21 (2012) pp. 475–576. doi:10.1017/S0962492912000062.
- [29] Francesco Leonetti, Francesco Siepe. Bounds for vector valued minimizers of some integral functionals. *Ricerche Mat.* 54 (2005) pp. 303–312 (2006).
- [30] Francesco Leonetti, Francesco Siepe. Maximum principle for vector valued minimizers. *J. Convex Anal.* 12 (2005) pp. 267–278.
- [31] Gwenael Mercier. Continuity results for TV-minimizers. *Indiana Univ. Math. J.* 67 (2018) pp. 1499–1545. doi:10.1512/iumj.2018.67.7393.
- [32] Leon Mirsky. A trace inequality of John von Neumann. *Monatsh. Math.* 79 (1975) pp. 303–306. doi:10.1007/BF01647331.
- [33] David Mumford, Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* 42 (1989) pp. 577–685. doi:10.1002/cpa.3160420503.
- [34] John von Neumann. *Collected works. Vol. IV: Continuous geometry and other topics.* Pergamon Press, Oxford-London-New York-Paris, (1962). General editor: A. H. Taub.
- [35] Franz Rellich. *Perturbation theory of eigenvalue problems.* Gordon and Breach Science Publishers, New York-London-Paris (1969). Assisted by J. Berkowitz, With a preface by Jacob T. Schwartz.
- [36] Leonid I. Rudin, Stanley Osher, Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D* 60 (1992) pp. 259–268. doi:10.1016/0167-2789(92)90242-F. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).
- [37] Roger Temam. *Problèmes mathématiques en plasticité*, vol. 12 of *Méthodes Mathématiques de l'Informatique [Mathematical Methods of Information Science]*. Gauthier-Villars, Montrouge (1983).
- [38] Tuomo Valkonen. The jump set under geometric regularization. Part 1: basic technique and first-order denoising. *SIAM J. Math. Anal.* 47 (2015) pp. 2587–2629. doi:10.1137/140976248.
- [39] Tuomo Valkonen. The jump set under geometric regularisation. Part 2: Higher-order approaches. *J. Math. Anal. Appl.* 453 (2017) pp. 1044–1085. doi:10.1016/j.jmaa.2017.04.037.
- [40] Singanallur V. Venkatakrishnan, Charles A. Bouman, Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing* (2013) pp. 945–948. doi:10.1109/GlobalSIP.2013.6737048.

- [41] Meiqin Wang, Liangliang Zhou, Jinhua Hu, Huiqun Cao, Danying Lin, Bin Yu, Junle Qu. Snapshot temporal compressive light-sheet fluorescence microscopy via deep denoising and total variation priors. *Opt. Lett.* 48 (2023) pp. 1144–1147. doi:10.1364/OL.475648.
- [42] Daniel Zoran, Yair Weiss. From learning models of natural image patches to whole image restoration. In *2011 International Conference on Computer Vision* (2011) pp. 479–486. doi:10.1109/ICCV.2011.6126278.