

## ANALYSIS AND APPLICATION OF A NON-LOCAL HESSIAN

JAN LELLMANN, KONSTANTINOS PAPAITSOROS, CAROLA SCHÖNLIEB AND DANIEL SPECTOR

**ABSTRACT.** In this work we introduce a formulation for a non-local Hessian that combines the ideas of *higher-order* and *non-local* regularization for image restoration, extending the idea of non-local gradients to higher-order derivatives. By carefully choosing the weights, the model allows to improve on the current state of the art higher-order method, Total Generalized Variation, with respect to overall quality and particularly close to jumps in the data. In the spirit of recent work by Brezis et al., our formulation also has analytic implications: for a suitable choice of weights, it can be shown to converge to classical second-order regularizers, and in fact allows a novel characterization of higher-order Sobolev and BV spaces.

**Keywords.** Non-local Hessian, Non-local Total Variation Regularization, Variational Methods, Fast Marching Method, Amoeba Filters

## 1. INTRODUCTION

The total variation model of image restoration due to Rudin, Osher and Fatemi [ROF92] is now classical, the problem of being given a noisy image  $g \in L^2(\Omega)$  on an open set  $\Omega \subseteq \mathbb{R}^2$  and selecting a restored image via minimization of the energy

$$E(u) := \int_{\Omega} (u - g)^2 dx + \alpha \text{TV}(u).$$

Here,  $\alpha > 0$  is a regularization parameter at our disposal and  $\text{TV}(u) := |Du|(\Omega)$  is the total variation of the measure that represents the distributional derivative of a function  $u \in \text{BV}(\Omega)$ , the space of functions of bounded variation. Among the known defects of the model is the staircasing effect, stemming from the use of the TV term in regularization. It is natural, then, to investigate the replacement of the total variation with another regularizer, for instance a higher-order term (see [Sch98, LLT03, LT06, HS06, CEP07, PS14] for the bounded Hessian framework, [CL97, BKP10, SST11] for infimal convolution and generalizations, and [LBL13] for anisotropic variants) or a non-local one (see, for example, the work of Buades, Coll and Morel [BCM05], Kinderman, Osher and Jones [KOJ05], and Gilboa and Osher [GO08]). In this work, we introduce and analyze a regularizer that is both higher-order and non-local, and utilize it in a model for image restoration. The heuristic interpretation of our model is based on a non-local Hessian, and our use of this nomenclature is justified through a rigorous localization analysis. Moreover, we give some numerical experiments to demonstrate that using our non-local Hessian, and by a suitable choice of weights, it is possible to derive specialized models that compete with current state of the art higher-order methods such as the Total Generalized Variation [BKP10].

**Background on non-local gradients.** In the first order setting, our analysis of non-local gradients and their associated energies find its origins in the 2001 paper of Bourgain, Brezis and Mironescu [BBM01]. In their paper, they had introduced energies of the form

$$(1.1) \quad F_n u := \int_{\Omega} \left( \int_{\Omega} \frac{|u(x) - u(y)|^{pq}}{|x - y|^{pq}} \rho_n(x - y) dx \right)^{\frac{1}{q}} dy,$$

where  $\Omega$  is a smooth bounded domain in  $\mathbb{R}^N$  and  $1 \leq p < \infty$ , and in the special case  $q = 1$ . Here, the functions  $\rho_n$  are radial mollifiers that are assumed to satisfy the following three properties for all  $n \in \mathbb{N}$ :

$$(1.2) \quad \rho_n(x) \geq 0,$$

$$(1.3) \quad \int_{\mathbb{R}^N} \rho_n(x) dx = 1,$$

$$(1.4) \quad \lim_{n \rightarrow \infty} \int_{|x| > \gamma} \rho_n(x) dx = 0, \quad \forall \gamma > 0.$$

The work [BBM01] connects the finiteness of the limit as  $n \rightarrow \infty$  of the functional (1.1) with the inclusion of a function  $u \in L^p(\Omega)$  in the Sobolev space  $W^{1,p}(\Omega)$  if  $p > 1$  or  $BV(\Omega)$  if  $p = 1$ . As in the beginning of the introduction, the space  $BV(\Omega)$  refers to the space of functions of bounded variation, and it is no coincidence that the two papers [BBM01, ROF92] utilize this energy space. Indeed, Gilboa and Osher [GO08] in 2008 independently introduced an energy similar to (1.1), terming it a non-local total variation, while the connection of the two and introduction of the parameter  $q$  is due to Leoni and Spector [LS11]. In particular, they show in [LS14] that for  $p = 1$  and  $q = 2$  the functionals (1.1)  $\Gamma$ -converge to a constant times the total variation. This result extends previous work by Ponce [Pon04b] in the case  $q = 1$  (see also the work of Aubert and Kornprobst [AK09] for an application of these results to image processing).

Gilboa and Osher [GO08] in fact introduced two forms of non-local total variations, and for our purposes here it will be useful to consider the second. This alternative involves introducing a non-local gradient operator, defined by

$$(1.5) \quad G_n u(x) := N \int_{\Omega} \frac{u(x) - u(y)}{|x - y|} \frac{x - y}{|x - y|} \rho_n(x - y) dy, \quad x \in \Omega,$$

for  $u \in C_c^1(\Omega)$ . Then, one defines the non-local total variation as the  $L^1$  norm of (1.5). The localization analysis of the non-local gradient (1.5) has been performed by Mengesha and Spector in [MS14], where a more general (and technical) distributional definition is utilized. Their first observation is that the definition of the non-local gradient via the Lebesgue integral (1.5) extends to spaces of weakly differentiable functions. In this regime they discuss the localization of (1.5). They prove that the non-local gradient converges to its local analogue  $Du$  in a topology that corresponds to the regularity of the underlying function  $u$ . As a result, they obtain yet another characterization of the spaces  $W^{1,p}(\Omega)$  and  $BV(\Omega)$ . Of notable interest for image processing purposes is their result on the  $\Gamma$ -convergence of the corresponding non-local total variation energies defined via non-local gradients of the form (1.5) to the local total variation.

**Background on higher-order regularization.** As the total variation and many more contributions in the image processing community have proven, a non-smooth regularization procedure indeed results in a nonlinear smoothing of the image, smoothing more in homogeneous areas of the image domain and preserving characteristic structures such as edges. In particular, the total variation regularizer is tuned towards the preservation of edges and performs very well if the reconstructed image is piecewise constant. The drawback of such a regularization procedure becomes apparent as soon as images or signals (in 1D) are considered which do not only consist of flat regions and jumps, but also possess slanted regions, i.e., piecewise affine parts. The artifact introduced by total variation regularization in this case is called staircasing [Rin00].

In [CL97] Chambolle and Lions propose a higher-order method by means of an infimal convolution of two convex regularizers. Here, a noisy image is decomposed into three parts  $g = u_1 + u_2 + n$  by solving

$$(1.6) \quad \min_{(u_1, u_2)} \frac{1}{2} \int_{\Omega} (u_1 + u_2 - g)^2 dx + \alpha \text{TV}(u_1) + \beta \text{TV}^2(u_2),$$

where  $\text{TV}^2(u_2) := D^2 u_2$  is the distributional Hessian of  $u_2$ . Then,  $u_1$  and  $u_2$  are the piecewise constant and piecewise affine parts of  $g$  respectively and  $n$  the noise (or texture). For recent modifications of this approach in the discrete setting see also [SS08, SST11]. Other approaches to combine first and second regularization originate, for instance, from Chan, Marquina, and Mulet [CMM01] who consider total variation minimisation together with weighted versions of the Laplacian, the Euler-astica functional [MM98, CKS02] which combines total variation regularization with curvature penalisation, and many more [LT06, LTC13, PS14, PSS13]. Recently Bredies et al. have proposed another interesting higher-order total variation model called Total Generalized Variation (TGV) [BKP10]. The TGV regularizer of order  $k$  is of the form

$$(1.7) \quad \text{TGV}_{\alpha}^k(u) = \sup \left\{ \int_{\Omega} u \operatorname{div}^k \xi \, dx : \xi \in C_c^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^N)), \|\operatorname{div}^l \xi\|_{\infty} \leq \alpha_l, l = 0, \dots, k-1 \right\},$$

where  $\operatorname{Sym}^k(\mathbb{R}^N)$  denotes the space of symmetric tensors of order  $k$  with arguments in  $\mathbb{R}^N$ , and  $\alpha = \{\alpha_l\}_{l=0}^{k-1}$  are fixed positive parameters. Its formulation for the solution of general inverse problems was given in [BV11].

The idea of pure bounded Hessian regularization is considered by Lysaker et al. [LLT03], Scherzer et al. [Sch98, HS06], Lefkimiatis et al. [LBU12] and Bergounioux and Piffet [BP10]. In these works the

considered model has the general form

$$\min_u \frac{1}{2} \int_{\Omega} (u - g)^2 dx + \alpha |D^2 u|(\Omega).$$

In [CEP07], Chan et al. use the squared  $L^2$  norm of the Laplacian as a regularizer also in combination with the  $H^{-1}$  norm in the data fitting term. Further, in [PS08] minimizers of functionals which are regularized by the total variation of the  $(l - 1)$ -th derivative, i.e.,

$$|D\nabla^{l-1}u|(\Omega),$$

are studied. Properties of such regularizers in terms of diffusion filters are further studied in [DWB09]. Therein, the authors consider the Euler-Lagrange equations corresponding to minimizers of functionals of the general type

$$\mathcal{J}(u) = \int_{\Omega} (u - g)^2 dx + \alpha \int_{\Omega} f \left( \sum_{|k|=p} |D^k u|^2 \right) dx,$$

for different non-quadratic functions  $f$ . There are also works on higher-order PDE methods for image regularization, e.g. [CS01, LLT03, BG04, BEG08, BHS09].

Confirmed by all of these works on higher-order total variation regularization the introduction of higher-order derivatives can have a positive effect on artifacts like the staircasing effect inherent to total variation regularization. In [DMFLM09], for instance, the authors analyse the positive effect of the Chan, Marquina, and Mulet model on the staircasing artifact.

**Higher-order non-local regularization.** From this perspective, the development of a non-local higher-order method promises to be interesting. One approach to a non-local higher-order method would be to approximate the gradients using non-local gradients as developed in [MS14]. For example, one can define a form of non-local Hessian as

$$(1.8) \quad G_n(\nabla u)(x) = N \int_{\Omega} \frac{\nabla u(x) - \nabla u(y)}{|x - y|} \otimes \frac{x - y}{|x - y|} \rho_n(x - y) dy,$$

and obtain some straightforward characterization of  $W^{2,p}(\mathbb{R}^N)$  and  $BV^2(\mathbb{R}^N)$ , the latter being the space of  $W^{1,1}(\mathbb{R}^N)$  functions whose second order distributional derivative is a bounded Radon measure. Here  $\otimes$  denotes the standard tensor multiplication of vectors. However, we find it advantageous to introduce and utilize a non-local Hessian that is derivative free, thus only requiring a function  $u$  to belong to an  $L^p$  space (instead of  $W^{1,p}$ ). In analogy with the theory of distributional derivatives, we first need a notion of non-local Hessian which can be applied to smooth functions. This definition, via a Lebesgue integral, is as follows.

**Definition 1.1.** *Suppose  $u \in C_c^2(\mathbb{R}^N)$ . Then we define the non-local Hessian as the Lebesgue integral*

$$(1.9) \quad H_n u(x) := \frac{N(N+2)}{2} \int_{\mathbb{R}^N} \frac{u(x+h) - 2u(x) + u(x-h)}{|h|^2} \left( h \otimes h - \frac{|h|^2}{N+2} I_N \right) \frac{\rho_n(h)}{|h|^2} dh, \quad x \in \mathbb{R}^N,$$

where here  $I_N$  is the  $N \times N$  identity matrix and  $\rho_n$  is a sequence satisfying (1.2)–(1.4).

We can then define the distributional non-local Hessian via its action on test functions.

**Definition 1.2.** *Suppose  $u \in L^p(\mathbb{R}^N)$ . Then we define the distributional non-local Hessian componentwise as*

$$(1.10) \quad \langle \mathfrak{H}_n^{ij} u, \varphi \rangle := \int_{\mathbb{R}^N} u H_n^{ij} \varphi dx,$$

for  $\varphi \in C_c^\infty(\mathbb{R}^N)$ .

A natural question is then whether these two notions agree. The following theorem shows that this is the case, provided the Lebesgue integral exists.

**Theorem 1.3** (Non-local Integration by Parts). *Suppose that  $u \in L^p(\mathbb{R}^N)$  for some  $1 \leq p < +\infty$  and  $\frac{|u(x+h)-2u(x)+u(x-h)|}{|h|^2} \rho_n(h) \in L^1(\mathbb{R}^N \times \mathbb{R}^N)$ . Then  $\mathfrak{H}_n u \in L^1(\mathbb{R}^N, \mathbb{R}^{N \times N})$  and for any  $\varphi \in C_c^2(\mathbb{R}^N)$  and  $i, j = 1 \dots N$ ,*

$$(1.11) \quad \langle \mathfrak{H}_n^{ij} u, \varphi \rangle = \int_{\mathbb{R}^N} H_n^{ij} u(x) \varphi(x) dx,$$

and therefore  $H_n u = \mathfrak{H}_n u$  almost everywhere as functions.

We will see in Section 3, Lemmas 3.1 and 3.2 that the Lebesgue integral even makes sense for  $u \in W^{2,p}(\mathbb{R}^N)$  or  $BV^2(\mathbb{R}^N)$ , and therefore the distributional definition  $\mathfrak{H}_n u$  coincides with the Lebesgue integral for these functions.

In the following, we undertake the rigorous analysis of the non-local Hessian, proving localization results in various topologies and characterizations of higher-order spaces of weakly differentiable functions. Our first result is the following theorem concerning the localization in the smooth case.

**Theorem 1.4.** *Suppose that  $u \in C_c^2(\mathbb{R}^N)$ . Then for any  $1 \leq p \leq +\infty$ ,*

$$H_n u \rightarrow \nabla^2 u, \quad \text{in } L^p(\mathbb{R}^N, \mathbb{R}^{N \times N}) \quad \text{as } n \rightarrow \infty.$$

When less smoothness is assumed on  $u$ , we have analogous convergence theorems, where the topology of convergence depends on the smoothness of  $u$ . When  $u \in W^{2,p}(\mathbb{R}^N)$ , we have the following.

**Theorem 1.5.** *Let  $1 \leq p < \infty$ . Then for every  $u \in W^{2,p}(\mathbb{R}^N)$  we have that*

$$H_n u \rightarrow \nabla^2 u \quad \text{in } L^p(\mathbb{R}^N, \mathbb{R}^{N \times N}) \quad \text{as } n \rightarrow \infty.$$

In the setting of  $BV^2(\mathbb{R}^N)$  (see Section 2 for a definition), we have the following theorem on the localization of the non-local Hessian.

**Theorem 1.6.** *Let  $u \in BV^2(\mathbb{R}^N)$ . Then*

$$\mu_n \rightarrow D^2 u, \quad \text{weakly}^*,$$

i.e., for every  $\phi \in C_0(\mathbb{R}^N, \mathbb{R}^{N \times N})$

$$(1.12) \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^N} H_n u(x) \cdot \phi(x) dx = \int_{\mathbb{R}^N} \phi(x) \cdot dD^2 u.$$

We have seen that the non-local Hessian is well-defined as a Lebesgue integral and localizes for spaces of weakly differentiable functions. In fact, it is sufficient to assume that  $u \in L^p(\mathbb{R}^N)$  is a function such that the distributions  $\mathfrak{H}_n u$  are in  $L^p(\mathbb{R}^N, \mathbb{R}^{N \times N})$  with a uniform bound of their  $L^p$  norms, in order to deduce that  $u \in W^{2,p}(\mathbb{R}^N)$  if  $1 < p < +\infty$  or  $u \in BV^2(\mathbb{R}^N)$  if  $p = 1$ . Precisely, we have the following theorems characterizing the second order Sobolev and BV spaces.

**Theorem 1.7.** *Let  $u \in L^p(\mathbb{R}^N)$  for some  $1 < p < \infty$ . Then*

$$(1.13) \quad u \in W^{2,p}(\mathbb{R}^N) \iff \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^N} |\mathfrak{H}_n u(x)|^p dx < \infty.$$

Let now  $u \in L^1(\mathbb{R}^N)$ . Then

$$(1.14) \quad u \in BV^2(\mathbb{R}^N) \iff \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^N} |\mathfrak{H}_n u(x)| dx < \infty.$$

While it is a natural extension of the first-order model and very well suited for a rigorous analysis, the definition of the non-local Hessian in (1.9) has some drawbacks in practice. In order to utilize the full potential of non-local models in image processing practice, it is natural to allow arbitrary, fixed (non-localizing) weights  $\sigma_x(h)$ . The difficulty of accommodating this in our current model comes from the fact that (1.9) has a built-in symmetry that associates triples of points rather than pairs which makes dealing with bounded domains more difficult and reduces the freedom in choosing the weights. Therefore we also propose an alternative formulation of the non-local Hessian that is closely related to (1.9) but allows more freedom and is more

amenable to numerical implementation. It is based on the observation that ultimately all finite-differences discretizations of the second-order derivatives stem from the requirement that they should be compatible with the Taylor expansion for smooth functions. More precisely, we can directly define the *implicit* non-local gradient  $G_u(x) \in \mathbb{R}^N$  and Hessian  $H_u(x) \in \text{Sym}(\mathbb{R}^{N \times N})$  that *best explain  $u$  around  $x$  in terms of a quadratic model*:

$$(1.15) \quad (G'_u(x), H'_u(x)) = \underset{G_u \in \mathbb{R}^N, H_u \in \text{Sym}(\mathbb{R}^{N \times N})}{\text{argmin}} \frac{1}{2} \int_{\Omega - \{x\}} \left( u(x+z) - u(x) - G_u^\top z - \frac{1}{2} z^\top H_u z \right)^2 \sigma_x(z) dz,$$

where  $\Omega \setminus \{x\} = \{y - x : y \in \Omega\}$ . While more complicated at first, this definition is equivalent to the explicit model (1.9) under a certain choice of the weights, see Theorem 4.2. As the objectives of the minimization problems (1.15) are quadratic, their solutions can be characterized by linear optimality conditions. Thus functionals based on the implicit non-local derivatives can be easily included in usual convex solvers by adding these conditions. Moreover, the weights  $\sigma_x(z)$  between any pair of points  $x$  and  $y = x + z$  can be chosen arbitrarily, without any restrictions on symmetry. These advantages come at the cost of being considerably harder to treat analytically, and we leave a rigorous definition and analysis of (1.15) for future work.

Of practical relevance is the question of how to choose the weights  $\sigma_x$  for a particular purpose. We show an example where the weights are used to construct a regularizer that both favours piecewise affine functions but also allows for jumps in the data. It is motivated by the recent discussion of ‘‘Amoeba’’ filters in [LDM07, WBV11, Wel12] which combine standard filters such as median filters with non-parametric structuring elements that are based on the data – that is in long thin objects they would extend along the structure and thus prevent smoothing perpendicular to the structure. In Amoeba filtering, the shape of a structuring element at a point is defined as a unit circle with respect to the geodesic distance on a manifold defined by the image itself. Motivated by their results we propose to extend their method to variational image processing by using the geodesic distance to set the weights  $\sigma_x$ . This allows to get a very close approximation to true piecewise affine regularization, in many cases improving on the results obtained using TGV.

Recently there has been another approach at defining a higher-order extension of non-local regularization for TGV-based energies [RBP14]. The authors start with the cascading formulation of TGV,

$$\text{TGV}^2(u) = \inf_{w: \Omega \rightarrow \mathbb{R}^N} \alpha_1 \int_{\Omega} |Du - w| + \alpha_0 \int_{\Omega} |Dw|,$$

which reduces the higher-order differential operators that appear in the definition of TGV to an infimal convolution of two terms involving only first-order derivatives. These can then be replaced by classical first-order non-local derivatives, and one obtains an energy of the form

$$\inf_{w: \Omega \rightarrow \mathbb{R}^N} \int_{\Omega} \int_{\Omega} \alpha_1(x, y) |u(x) - u(y) - w(x)^\top (x - y)| dy dx + \sum_{i=1}^2 \int_{\Omega} \int_{\Omega} \alpha_0(x, y) |w^i(x) - w^i(y)| dy dx.$$

Comparing this to (1.15), we see that the left integral is similar to our implicit model for first-order derivatives with  $w$  taking the role of  $G_u$ , but with the square in (1.15) replaced by a  $L^1$ -type norm and added to the energy as a penalty term rather than strictly enforcing optimality.

Finally, let us mention an important localization result from the perspective of variational image processing, the following theorem asserting the  $\Gamma$ -convergence [DM93, Bra02] of the non-local Hessian energies to the energy of the Hessian.

**Theorem 1.8.** *Assume that  $\rho_n$  satisfy (1.2), (1.3) and (1.4). Then*

$$\Gamma_{L^1(\mathbb{R}^N)}^- \lim_{n \rightarrow \infty} \int_{\mathbb{R}^N} |\mathfrak{H}_n u| dx = |D^2 u|(\mathbb{R}^N),$$

where the  $\Gamma$ -limit is taken with respect to the strong convergence  $u_n \rightarrow u$  in  $L^1(\mathbb{R}^N)$ .

The relevance of this theorem in the context of variational problems comes from the fact that  $\Gamma$ -convergence of the *objective functions* of a sequence of minimization problems implies convergence of the minimizers in a suitable topology [Bra02, Ch. 1.5]. Therefore Theorem 1.8 guarantees that under a suitable choice of weights, the solutions of a class of *non-local* Hessian-based problems converges to the solution of the *local* Hessian-regularized problem, and thus our notion of ‘‘non-local Hessian’’ is justified.

**Organization of the paper.** The paper is organized as follows: In Section 2 we recall some preliminary notions and we fix our notation. Section 3 deals with the analysis of the non-local Hessian functional (1.9). After a justification of the introduction of its distributional form, we proceed in Section 3.1 with the localization of (1.9) to the classical Hessian for smooth functions  $u$ . The localization of (1.9) to its classical analogue for  $W^{2,p}(\mathbb{R}^N)$  and  $BV^2(\mathbb{R}^N)$  functions is shown in Sections 3.2 and 3.3 respectively. In Section 3.4 we provide the non-local characterizations of the spaces  $W^{2,p}(\mathbb{R}^N)$  and  $BV^2(\mathbb{R}^N)$  in the spirit of [BBM01]. The  $\Gamma$ -convergence result, Theorem 1.8, is proved in Section 3.5.

The introduction of the implicit formulation of non-local Hessian (1.15), along with its connection to the explicit one, is presented in Section 4.1. In Section 4.2 we describe how we choose the weights  $\sigma_x$  in (1.15) in order to achieve jump preservation in the restored images. Finally, in Section 4.3 we present our numerical results, comparing our method with TGV.

## 2. PRELIMINARIES AND NOTATION

For the reader's convenience we recall here some important notions that we are going to use in the following sections and we also fix some notation.

As far as our notation is concerned, whenever a function space has two arguments, the first always denotes the function's domain while the second denotes its range. Whenever the range is omitted, it is assumed that the functions are real valued.

We use  $dx, dy, dz$  for various integrations with respect Lebesgue measure on  $\mathbb{R}^N$ , while in Section 3 we will have occasion to use the more succinct notation  $\mathcal{L}^{N^2}$  to denote integration with respect to the Lebesgue measure in the product space  $\mathbb{R}^N \times \mathbb{R}^N$ .

The reader should not confuse the different forms of the letter "H". We denote by  $\mathcal{H}$  the one-dimensional Hausdorff measure ( $\mathcal{H}^N$  for the N-dimensional), while  $H$  denotes the non-local Hessian when this is a function. As we have already seen,  $\mathfrak{H}$  denotes the distributional form of the non-local Hessian.

It is also very convenient to introduce the following notation:

$$d^2u(x, y) := u(y) - 2u(x) + u(x + (x - y)),$$

which can be interpreted a discrete second order differential operator in  $x$  at the direction  $x - y$ .

We denote by  $|\cdot|$  the Euclidean norm (vectors) and Frobenius norm (matrices).

As usual, we denote by  $BV(\Omega)$  the space of functions of bounded variation defined on an open  $\Omega \subseteq \mathbb{R}^N$ . This space consists of all real valued functions  $u \in L^1(\Omega)$  whose distributional derivative  $Du$  can be represented by a finite Radon measure. The total variation  $TV(u)$  of a function  $u \in BV(\Omega)$ , is defined to be the total variation of the measure  $Du$ , i.e.,  $TV(u) := |Du|(\Omega)$ . The definition is similar for vector valued functions. We refer the reader to [AFP00] for a full account of the theory of BV functions.

We denote by  $BV^2(\Omega)$  the space of functions of bounded Hessian. These are all the functions that belong to the Sobolev space  $W^{1,1}(\Omega)$  such that  $\nabla u$  is an  $\mathbb{R}^N$ -valued BV function, i.e.,  $\nabla u \in BV(\Omega, \mathbb{R}^N)$  and we set  $D^2u := D(\nabla u)$ . We refer the reader to [Dem85, BP10, PS14] for more information about this space. Let us however state a theorem that will be useful for our purposes. It is the analogue result to the strict approximation by smooth functions for the classical BV case, see [AFP00].

**Theorem 2.1** ( $BV^2$  strict approximation by smooth functions, [Dem85]). *Let  $\Omega \subseteq \mathbb{R}^N$  open and  $u \in BV^2(\Omega)$ . Then there exists a sequence  $(u_n)_{n \in \mathbb{N}} \in W^{2,1}(\Omega) \cap C^\infty(\Omega)$  that converges to  $u$  strictly in  $BV^2(\Omega)$ , that is to say*

$$u_n \rightarrow u, \quad \text{in } L^1(\Omega) \quad \text{and} \quad |D^2u_n|(\Omega) \rightarrow |D^2u|(\Omega), \quad \text{as } n \rightarrow \infty.$$

We recall also the two basic notions of convergence regarding finite Radon measures. We note that  $\mathcal{M}(\Omega, \mathbb{R}^\ell)$  denotes the space of  $\mathbb{R}^\ell$ -valued finite Radon measures in  $\Omega$ . If  $(\mu_n)_{n \in \mathbb{N}}$  and  $\mu$  are real valued finite Radon measures defined on an open  $\Omega \subseteq \mathbb{R}^N$  we say that the sequence  $\mu_n$  converges weakly\* to  $\mu$  if for all  $\phi \in C_0(\Omega)$  we have  $\int_\Omega \phi d\mu_n \rightarrow \int_\Omega \phi d\mu$  as  $n$  goes to infinity. Here  $\phi \in C_0(\Omega)$  means that  $\phi$  is continuous on  $\Omega$  and that for every  $\epsilon > 0$  there exists a compact set  $K \subset \Omega$  such that  $\sup_{x \in \Omega \setminus K} |\phi(x)| \leq \epsilon$ . Note that from the Riesz representation theorem the dual space  $(C_0(\Omega, \mathbb{R}^\ell), \|\cdot\|_\infty)^*$  can be identified with  $\mathcal{M}(\Omega, \mathbb{R}^\ell)$ . We say that the convergence is strict if in addition to that we also have that  $|\mu_n|(\Omega) \rightarrow |\mu|(\Omega)$ , i.e., the total variations of  $\mu_n$  converge to the total variation of  $\mu$ . The definition is similar for vector and matrix valued measures with all the operations regarded component-wise.

We now remind the reader about some basic facts concerning  $\Gamma$ -convergence. Let  $(X, d)$  be a metric space and  $F, F_n : X \rightarrow \mathbb{R} \cup \{+\infty\}$  for all  $n \in \mathbb{N}$ . We say that the sequence of functionals  $F_n$   $\Gamma$ -converges to  $F$  at  $x \in X$  in the topology of  $X$  and we write  $\Gamma_X^- \text{-lim}_{n \rightarrow \infty} F_n(x) = F(x)$  if the following two conditions hold

- (1) For every sequence  $(x_n)_{n \in \mathbb{N}}$  converging to  $x$  in  $(X, d)$  we have

$$F(x) \leq \liminf_{n \rightarrow \infty} F_n(x_n).$$

- (2) There exists a sequence  $(x_n)_{n \in \mathbb{N}}$  converging to  $x$  in  $(X, d)$  such that

$$F(x) \geq \limsup_{n \rightarrow \infty} F_n(x_n).$$

It can be proved that  $\Gamma_X^- \text{-lim}_{n \rightarrow \infty} F_n(x) = F(x)$  if the  $\Gamma$ -lower and  $\Gamma$ -upper limits of  $F_n$  at  $x$ , denoted by  $\Gamma_X^- \text{-lim inf}_{n \rightarrow \infty} F_n(x)$  and  $\Gamma_X^- \text{-lim sup}_{n \rightarrow \infty} F_n(x)$  respectively, are equal to  $F(x)$  where

$$\begin{aligned} \Gamma_X^- \text{-lim inf}_{n \rightarrow \infty} F_n(x) &= \min\{\liminf_{n \rightarrow \infty} F_n(x_n) : x_n \rightarrow x \text{ in } (X, d)\}, \\ \Gamma_X^- \text{-lim sup}_{n \rightarrow \infty} F_n(x) &= \min\{\limsup_{n \rightarrow \infty} F_n(x_n) : x_n \rightarrow x \text{ in } (X, d)\}. \end{aligned}$$

Finally, if  $F : X \rightarrow \mathbb{R} \cup \{+\infty\}$  we denote by  $sc_X^- F$  the lower semicontinuous envelope of  $F$ , i.e., the greatest lower semicontinuous function majorized by  $F$ . We refer the reader to [DM93, Bra02] for further details regarding  $\Gamma$ -convergence and lower semicontinuous envelopes.

### 3. ANALYSIS OF THE NON-LOCAL HESSIAN

The precise form we have chosen for the non-local Hessian can be derived from the model case of non-local gradients - the fractional gradient - which has been developed in [SS14]. Here we prove several results analogous to the first order case, as in [MS14], for the generalizations involving generic radial weights that satisfy (1.2)–(1.4). Of primary importance is to first establish that the distributional non-local Hessian defined by equation (1.10) is, in fact, a distribution. Here we observe that if  $u \in L^1(\mathbb{R}^N)$ , then

$$|\langle \mathfrak{H}_n u, \varphi \rangle| \leq C \|u\|_{L^1(\mathbb{R}^N)} \|\nabla^2 \varphi\|_{L^\infty(\mathbb{R}^N)},$$

so that  $\mathfrak{H}_n u$  is a distribution. Also observe that if  $u \in L^p(\mathbb{R}^N)$  for some  $1 < p < \infty$ , then from the estimate (3.4) below together with the fact that  $\varphi$  is of compact support we have

$$|\langle \mathfrak{H}_n u, \varphi \rangle| \leq C \|u\|_{L^p(\mathbb{R}^N)} \|\nabla^2 \varphi\|_{L^q(\mathbb{R}^N, \mathbb{R}^{N \times N})} \leq C \|u\|_{L^p(\mathbb{R}^N)} \|\nabla^2 \varphi\|_{L^\infty(\mathbb{R}^N)},$$

where  $1/p + 1/q = 1$  and thus  $\mathfrak{H}_n u$  is indeed again a distribution. One observes that the definition is in analogy to the theory of Sobolev spaces, where weak derivatives are defined in terms of the integration by parts formula. Because the Hessian is composed of two derivatives, we observe that there is no change in sign in the definition, preserving some symmetry that will be useful for us in the sequel.

The second important item to address is the agreement of the distributional non-local Hessian with the non-local Hessian. The necessary and sufficient condition is the existence of the latter, which is the assertion of Theorem 1.3. We now substantiate this assertion.

*Proof of Theorem 1.3.* Let  $1 \leq p < +\infty$  and suppose  $u \in L^p(\mathbb{R}^N)$  and that  $H_n u$  is well-defined as a Lebesgue integral. Let  $\varphi \in C_c^2(\mathbb{R}^N)$ , and fix  $i, j \in \{1, \dots, N\}$ . Then it is a consequence of Fubini's theorem and Lebesgue's dominated convergence theorem that

$$\begin{aligned} & \int_{\mathbb{R}^N} H_n^{ij} u(x) \varphi(x) dx \\ &= \frac{N(N+2)}{2} \lim_{\epsilon \rightarrow 0} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N \setminus B(x, \epsilon)} \frac{d^2 u(x, y) \left( (x_i - y_i) \otimes (x_j - y_j) - \frac{|x-y|^2}{N+2} I_N \right)}{|x-y|^2 |x-y|^2} \rho(x-y) \varphi(x) dy dx \\ &= \frac{N(N+2)}{2} \lim_{\epsilon \rightarrow 0} \int_{d_\epsilon^N} \frac{d^2 u(x, y) \left( (x_i - y_i) \otimes (x_j - y_j) - \frac{|x-y|^2}{N+2} I_N \right)}{|x-y|^2 |x-y|^2} \rho(x-y) \varphi(x) d(\mathcal{L}^N)^2(x, y), \end{aligned}$$

where  $d_\epsilon^N := \mathbb{R}^N \times \mathbb{R}^N \setminus \{|x - y| < \epsilon\}$ . Similarly we have

$$\begin{aligned} & \int_{\mathbb{R}^N} u(x) H_n^{ij} \varphi(x) dx \\ &= \frac{N(N+2)}{2} \lim_{\epsilon \rightarrow 0} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N \setminus B(x, \epsilon)} u(x) \frac{d^2 \varphi(x, y)}{|x - y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x - y|^2}{N+2}}{|x - y|^2} \rho(x - y) dy dx \\ &= \frac{N(N+2)}{2} \lim_{\epsilon \rightarrow 0} \int_{d_\epsilon^N} u(x) \frac{d^2 \varphi(x, y)}{|x - y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x - y|^2}{N+2}}{|x - y|^2} \rho(x - y) d(\mathcal{L}^N)^2(x, y), \end{aligned}$$

where, for notational convenience, we used the standard convention

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Thus, it suffices to show that for every  $i, j$  and  $\epsilon > 0$  we have

$$(3.1) \quad \begin{aligned} & \int_{d_\epsilon^N} \frac{d^2 u(x, y)}{|x - y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x - y|^2}{N+2}}{|x - y|^2} \rho(x - y) \varphi(x) d(\mathcal{L}^N)^2(x, y) = \\ & \int_{d_\epsilon^N} u(x) \frac{d^2 \varphi(x, y)}{|x - y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x - y|^2}{N+2}}{|x - y|^2} \rho(x - y) d(\mathcal{L}^N)^2(x, y). \end{aligned}$$

In order to show (3.1), it suffices to prove

$$(3.2) \quad \begin{aligned} & \int_{d_\epsilon^N} \frac{u(y) \varphi(x)}{|x - y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x - y|^2}{N+2}}{|x - y|^2} \rho(x - y) d(\mathcal{L}^N)^2(x, y) = \\ & \int_{d_\epsilon^N} \frac{u(x) \varphi(y)}{|x - y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x - y|^2}{N+2}}{|x - y|^2} \rho(x - y) d(\mathcal{L}^N)^2(x, y), \end{aligned}$$

and

$$(3.3) \quad \begin{aligned} & \int_{d_\epsilon^N} \frac{u(x + (x - y)) \varphi(x)}{|x - y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x - y|^2}{N+2}}{|x - y|^2} \rho(x - y) d(\mathcal{L}^N)^2(x, y) = \\ & \int_{d_\epsilon^N} \frac{u(x) \varphi(x + (x - y))}{|x - y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x - y|^2}{N+2}}{|x - y|^2} \rho(x - y) d(\mathcal{L}^N)^2(x, y). \end{aligned}$$

Equation (3.2) can be easily showed by alternating  $x$  and  $y$  and using the symmetry of the domain. Finally equation (3.3) can be proved by employing the substitution  $u = 2x - y$ ,  $v = 3x - 2y$ , noting that  $x - y = v - u$  and that the determinant of the Jacobian of this substitution is  $-1$ .

Having established that the notion of distributional non-local Hessian and non-local Hessian agree whenever the latter exists, it is a natural question to ask when this is the case. It is a simple calculation to verify that the Lebesgue integral (1.9) exists whenever  $u \in C_c^2(\mathbb{R}^N)$ . The following several theorems, whose proofs we also defer until after the analysis in the smooth case, show that this is also the case for functions in the spaces  $W^{2,p}(\mathbb{R}^N)$  and  $BV^2(\mathbb{R}^N)$ .

**Lemma 3.1.** *Suppose that  $u \in W^{2,p}(\mathbb{R}^N)$ , where  $1 \leq p < \infty$ . Then  $H_n u$  is well-defined as a Lebesgue integral,  $H_n u \in L^p(\mathbb{R}^N, \mathbb{R}^{N \times N})$ , and*

$$(3.4) \quad \int_{\mathbb{R}^N} |H_n u(x)|^p dx \leq M \|\nabla^2 u\|_{L^p(\mathbb{R}^N)}^p,$$

where the constant  $M$  depends only on  $N$  and  $p$ .

**Lemma 3.2.** *Suppose that  $u \in BV^2(\mathbb{R}^N)$ . Then  $H_n u \in L^1(\mathbb{R}^N, \mathbb{R}^{N \times N})$  with*

$$(3.5) \quad \int_{\mathbb{R}^N} |H_n u(x)| dx \leq M |D^2 u|(\mathbb{R}^N),$$

where the constant  $M$  depends only on  $N$ .

We finally point out that all our initial analysis is done in  $\mathbb{R}^N$ . Defining (1.9) in a general  $\Omega \subseteq \mathbb{R}^N$  is problematic as the term  $x + (x - y)$  does not necessarily belong to  $\Omega$  whenever  $x, y \in \Omega$ .

**3.1. Localization – Smooth case.** We are now ready to prove the localization of  $H_n u$  to  $\nabla^2 u$  for smooth functions.

*Proof of Theorem 1.4.*

**Case  $1 \leq p < +\infty$**

Let us assume that we have shown the case  $p = +\infty$ . Then we must show that the uniform convergence  $H_n v \rightarrow \nabla^2 v$  for  $v \in C_c^2(\mathbb{R}^N)$ , implies convergence in  $L^p(\mathbb{R}^N, \mathbb{R}^{N \times N})$  for any  $1 \leq p < +\infty$ . We claim that this will follow from the following uniform estimate on the *tails* of the nonlocal Hessian. Suppose  $\text{supp } v \subset B(0, R)$ , where  $\text{supp } v$  denotes the support of  $v$ . Then for any  $1 \leq p < +\infty$  and  $\epsilon > 0$  there exists a  $L = L(\epsilon, p) > 1$  such that

$$(3.6) \quad \sup_n \int_{B(0, LR)^c} |H_n v(x)|^p dx \leq \epsilon.$$

If this were the case, we would estimate the  $L^p$ -convergence as follows

$$\int_{\mathbb{R}^N} |H_n v(x) - \nabla^2 v(x)|^p dx = \int_{B(0, LR)} |H_n v(x) - \nabla^2 v(x)|^p dx + \int_{B(0, LR)^c} |H_n v(x)|^p dx,$$

from which equation (3.6) implies

$$\limsup_{n \rightarrow \infty} \int_{\mathbb{R}^N} |H_n v(x) - \nabla^2 v(x)|^p dx \leq \limsup_{n \rightarrow \infty} \int_{B(0, LR)} |H_n v(x) - \nabla^2 v(x)|^p dx + \epsilon.$$

The conclusion then follows, since the first term vanishes from the uniform convergence assumed, after which  $\epsilon > 0$  is arbitrary. We will therefore show the estimate (3.6). We have, by Jensen's inequality with respect to the measure  $\rho_n$ , which has  $\int_{\mathbb{R}^N} \rho_n(x) dx = 1$ , that

$$\begin{aligned} \int_{B(0, LR)^c} |H_n v(x)|^p dx &\leq \int_{B(0, LR)^c} \int_{\mathbb{R}^N} \frac{|v(y) - 2v(x) + v(x + x - y)|^p}{|x - y|^{2p}} \rho_n(x - y) dy dx \\ &= \int_{B(0, LR)^c} \int_{y \in B(0, R)} \frac{|v(y)|^p}{|x - y|^{2p}} \rho_n(x - y) dy dx \\ &\quad + \int_{B(0, LR)^c} \int_{x + x - y \in B(0, R)} \frac{|v(x + x - y)|^p}{|x - y|^{2p}} \rho_n(x - y) dy dx. \end{aligned}$$

Letting  $z = x + x - y$  (which means that  $x - y = z - x$ ), we obtain

$$\int_{B(0, LR)^c} \int_{x + x - y \in B(0, R)} \frac{|v(x + x - y)|^p}{|x - y|^{2p}} \rho_n(x - y) dy dx = \int_{B(0, LR)^c} \int_{z \in B(0, R)} \frac{|v(z)|^p}{|z - x|^{2p}} \rho_n(z - x) dz dx,$$

and therefore by symmetry of  $\rho_n$  we have

$$\begin{aligned} \int_{B(0, LR)^c} |H_n v(x)|^p dx &\leq 2 \int_{B(0, LR)^c} \int_{y \in B(0, R)} \frac{|v(y)|^p}{|x - y|^{2p}} \rho_n(x - y) dy dx \\ &\leq \frac{2}{(L - 1)^{2p}} \int_{B(0, LR)^c} \int_{y \in B(0, R)} |v(y)|^p \rho_n(x - y) dy dx \\ &\leq \frac{2}{(L - 1)^{2p}} \|\rho_n\|_{L^1(\mathbb{R}^N)} \|v\|_{L^p(\mathbb{R}^N)}^p. \end{aligned}$$

Again using  $\int_{\mathbb{R}^N} \rho_n(x) dx = 1$ , the claim, and therefore the case  $1 \leq p < +\infty$ , then follows by choosing  $L$  sufficiently large.

**Case  $p = +\infty$**

It therefore remains to show that the convergence in  $L^\infty(\mathbb{R}^N, \mathbb{R}^{N \times N})$  is true. Precisely, we will show that

$$|H_n u - \nabla^2 u| \rightarrow 0, \quad \text{uniformly,}$$

for which it suffices to prove the convergence component wise, i.e.,  $\left| (H_n u - \nabla^2 u)_{(i_0, j_0)} \right| \rightarrow 0$  by considering two cases  $i_0 \neq j_0$  and  $i_0 = j_0$ .

**Subcase  $i_0 \neq j_0$  :**

Let us first observe that Proposition 4.1 in the Appendix and the assumption that  $\int_{\mathbb{R}^N} \rho_n(x) dx = 1$  for all  $n \in \mathbb{N}$  can be used to deduce that

$$(3.7) \quad \begin{aligned} \int_{\mathbb{R}^N} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho_n(z) dz &= \int_0^\infty \rho_n(t) t^{N-1} dt \int_{\mathcal{S}^{N-1}} \nu_{i_0}^2 \nu_{j_0}^2 d\mathcal{H}^{N-1}(x) \\ &= \frac{1}{N(N+2)} \cdot \begin{cases} 1, & i_0 \neq j_0, \\ 3, & i_0 = j_0. \end{cases} \end{aligned}$$

Therefore, in the case that  $i_0 \neq j_0$ , we have

$$\begin{aligned} & \left| (H_n u - \nabla^2 u)_{(i_0, j_0)} \right| \\ &= \frac{N(N+2)}{2} \left| \int_{\mathbb{R}^N} \frac{d^2 u(x, y)}{|x-y|^2} \frac{(x_{i_0} - y_{i_0})(x_{j_0} - y_{j_0})}{|x-y|^2} \rho_n(x-y) dy - 2 \frac{\partial u}{\partial x_{i_0} \partial x_{j_0}}(x) \int_{\mathbb{R}^N} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho_n(z) dz \right|. \end{aligned}$$

However, again utilizing the radial symmetry of  $\rho_n$ , we have that the following integrals vanish:

$$\begin{aligned} \mathbf{i}_0 = \mathbf{j}_0 &: \int_{\mathbb{R}^N} \frac{z_i z_{j_0}^3}{|z|^4} \rho_n(z) dz = 0, \quad \text{for } i \neq j_0, \\ \mathbf{i}_0 = \mathbf{j}_0 &: \int_{\mathbb{R}^N} \frac{z_i z_j z_{j_0}^2}{|z|^4} \rho_n(z) dz = 0, \quad \text{for } i \neq j_0, j \neq j_0, i \neq j, \\ \mathbf{i}_0 \neq \mathbf{j}_0 &: \int_{\mathbb{R}^N} \frac{z_i z_j z_{i_0} z_{j_0}}{|z|^4} \rho_n(z) dz = 0, \quad \text{for } i \neq i_0, j \neq j_0, \end{aligned}$$

which implies that

$$\sum_{i,j=1}^N \frac{\partial u}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^N} \frac{z_i z_{i_0} z_j z_{j_0}}{|z|^4} \rho_n(z) dz = 2 \frac{\partial u}{\partial x_{i_0} \partial x_{j_0}}(x) \int_{\mathbb{R}^N} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho_n(z) dz.$$

Thus, introducing these factors of zero, and writing in a more compact way, we have that

$$\begin{aligned} & \left| (H_n u - \nabla^2 u)_{(i_0, j_0)} \right| \\ &= \frac{N(N+2)}{2} \left| \int_{\mathbb{R}^N} \frac{d^2 u(x, y) - (x-y)^T \nabla^2 u(x) (x-y)}{|x-y|^2} \frac{(x_{i_0} - y_{i_0})(x_{j_0} - y_{j_0})}{|x-y|^2} \rho_n(x-y) dy \right|. \end{aligned}$$

We want to show that the right hand side tends to zero as  $n \rightarrow \infty$ , and therefore we define now the following quantity for  $\delta > 0$ :

$$(3.8) \quad Q_\delta u(x) = \left| \int_{B(x, \delta)} \frac{d^2 u(x, y) - (x-y)^T \nabla^2 u(x) (x-y)}{|x-y|^2} \frac{(x_{i_0} - y_{i_0})(x_{j_0} - y_{j_0})}{|x-y|^2} \rho_n(x-y) dy \right|.$$

We then claim that we can make  $Q_\delta u(x)$  as small as we want, independently of  $x$  and  $n$ , by choosing sufficiently small  $\delta > 0$ . If this is the case, then the case  $i_0 \neq j_0$  would be completed, since we would then have that

$$\begin{aligned} \left| (H_n u - \nabla^2 u)_{(i_0, j_0)} \right| &\leq Q_\delta(x) + \frac{N(N+2)}{2} \int_{|z| \geq \delta} \frac{|u(x+z) - 2u(x) + u(x-z)|}{|z|^2} \rho_n(z) dz \\ &\quad + \frac{N(N+2)}{2} |\nabla^2 u(x)| \int_{|z| \geq \delta} \rho_n(z) dz \\ &\leq \frac{N(N+2)}{2} \epsilon + \frac{N(N+2)}{2} \left( \frac{4\|u\|_\infty}{\delta^2} + \|\nabla^2 u\|_\infty \right) \int_{|z| \geq \delta} \rho_n(z) dz \\ &< N(N+2)\epsilon, \end{aligned}$$

for  $n$  large enough, and the result follows from sending  $\epsilon \rightarrow 0$ .

We therefore proceed to make estimates for (3.8). Since we have assumed  $u \in C_c^2(\mathbb{R}^N)$ , we have that given  $\epsilon > 0$ , there is a  $\delta > 0$  such that for every  $i, j = 1, \dots, N$  we have

$$\left| \frac{\partial u}{\partial x_i \partial x_j}(x) - \frac{\partial u}{\partial x_i \partial x_j}(y) \right| < \epsilon, \quad \text{whenever } |x - y| < \delta.$$

Using (3.11) we can estimate

$$(3.9) \quad Q_\delta u(x) = \left| \int_{B(x, \delta)} \frac{d^2 u(x, y) - (x - y)^T \nabla^2 u(x)(x - y)}{|x - y|^2} \frac{(x_{i_0} - y_{i_0})(x_{j_0} - y_{j_0})}{|x - y|^2} \rho_n(x - y) dy \right|$$

$$(3.10) \quad = \left| \int_{B(x, \delta)} \frac{(x - y)^T \left( \int_0^1 \int_0^1 \nabla^2 u(x + (s + t - 1)(y - x)) - \nabla^2 u(x) ds dt \right) (x - y)}{|x - y|^2} \right. \\ \left. \times \frac{((x_{i_0} - y_{i_0})(x_{j_0} - y_{j_0})) \rho_n(x - y) dy}{|x - y|^2} \right| \\ \leq N \int_{B(x, \delta)} \frac{|x - y| \epsilon |x - y|}{|x - y|^2} \frac{|x_{i_0} - y_{i_0}| |x_{j_0} - y_{j_0}|}{|x - y|^2} \rho_n(x - y) dy \\ \leq \epsilon N.$$

Here, we have used the mean value theorem for scalar and vector valued functions to write

$$(3.11) \quad d^2 u(x, y) = (x - y)^T \left( \int_0^1 \int_0^1 \nabla^2 u(x + (t + s - 1)(y - x)) ds dt \right) (x - y),$$

and the fact that  $\int_{\mathbb{R}^N} \rho_n(x) dx = 1$  for all  $n \in \mathbb{N}$ . This completes the proof in the case  $i_0 \neq j_0$ .

**Subcase  $i_0 = j_0$  :**

Let us record several observations before we proceed with this case. In fact, the same argument shows that for a single  $i \in \{1, \dots, N\}$

$$(3.12) \quad I_i^n(x) := \left| \int_{\mathbb{R}^N} \frac{d^2 u(x, y) - (x - y)^T \nabla^2 u(x)(x - y)}{|x - y|^2} \frac{(x_i - y_i)^2}{|x - y|^2} \rho_n(x - y) dy \right| \rightarrow 0$$

uniformly in  $x$  as  $n \rightarrow \infty$ , and therefore by summing in  $i$  we deduce that

$$(3.13) \quad \left| \int_{\mathbb{R}^N} \frac{d^2 u(x, y) - (x - y)^T \nabla^2 u(x)(x - y)}{|x - y|^2} \rho_n(x - y) dy \right| \rightarrow 0.$$

Moreover, we observe that the same formula from Proposition 4.1 and cancellation of odd powers implies

$$\int_{\mathbb{R}^N} \frac{(x - y)^T \nabla^2 u(x)(x - y)(x_{i_0} - y_{i_0})^2}{|x - y|^4} \rho_n(x - y) dy = \sum_{j=1}^N \frac{\partial^2 u}{\partial x_j^2}(x) \int_{\mathbb{R}^N} \frac{z_j^2 z_{i_0}^2}{|z|^4} \rho_n(z) dz \\ = \frac{1}{N(N+2)} \Delta u + \frac{2}{3} \frac{\partial^2 u}{\partial x_{i_0}^2}(x) \int_{\mathbb{R}^N} \frac{z_{i_0}^4}{|z|^4} \rho_n(z) dz \\ = \frac{2}{N(N+2)} \left( \frac{1}{2} \Delta u + \frac{\partial^2 u}{\partial x_{i_0}^2}(x) \right),$$

while we also have that

$$\int_{\mathbb{R}^N} \frac{(x - y)^T \nabla^2 u(x)(x - y)}{|x - y|^2} \rho_n(x - y) dy = \sum_{j=1}^N \frac{\partial^2 u}{\partial x_j^2}(x) \int_{\mathbb{R}^N} \frac{z_j^2}{|z|^2} \rho_n(z) dz \\ = \frac{1}{N} \Delta u(x).$$

Thus, we can estimate

$$\left| (H_n u - \nabla^2 u)_{(i_0, i_0)} \right| \leq I_{i_0}^n(x) + \left| \frac{N}{2} \int_{\mathbb{R}^N} \frac{d^2 u(x, y)}{|x - y|^2} \rho_n(x - y) dy - \int_{\mathbb{R}^N} \frac{\Delta u(x)}{2} \rho_n(x - y) dy \right| \\ = I_{i_0}^n(x) + \left| \frac{N}{2} \int_{\mathbb{R}^N} \frac{d^2 u(x, y) - (x - y)^T \nabla^2 u(x)(x - y)}{|x - y|^2} \rho_n(x - y) dy \right|,$$

and the proof is completed by invoking the convergences established in equations (3.12) and (3.13).

**3.2. Localization –  $W^{2,p}(\mathbb{R}^N)$  case.** The objective of this section is to show that if  $u \in W^{2,p}(\mathbb{R}^N)$ ,  $1 \leq p < \infty$  then the non-local Hessian  $H_n u$  converges to  $\nabla^2 u$  in  $L^p$ . The first step is show that indeed in that case  $H_n u$  is indeed an  $L^p$  function. This follows from Lemma 3.1 which we prove next.

*Proof of Lemma 3.1.* Let us begin by making estimates for a function  $v \in C^\infty(\mathbb{R}^N) \cap W^{2,p}(\mathbb{R}^N)$ . From the definition of the non-local Hessian, and utilizing Jensen's inequality, equation (3.11) as well as Fubini's theorem we have the following successive estimates (the constant is always denoted with  $M(N, p)$ ):

$$\begin{aligned}
(3.14) \quad & \int_{\mathbb{R}^N} |H_n v(x)|_{\mathbb{R}^N \times \mathbb{R}^N}^p dx \\
&= \left( \frac{N(N+2)}{2} \right)^p \int_{\mathbb{R}^N} \left| \int_{\mathbb{R}^N} \frac{d^2 v(x, y)}{|x-y|^2} \frac{\left( (x-y) \otimes (x-y) - \frac{|x-y|^2}{N+2} I_N \right)}{|x-y|^2} \rho_n(x-y) dy \right|^p dx \\
&\leq M(N, p) \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{|d^2 v(x, y)|}{|x-y|^2} \rho_n(x-y) dy \right)^p dx \\
(3.15) \quad &\leq M(N, p) \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{|d^2 v(x, y)|^p}{|x-y|^{2p}} \rho_n(x-y) dy \right) \left( \underbrace{\int_{\mathbb{R}^N} \rho_n(x-y) dy}_{=1} \right)^{p/p'} dx \\
&\leq M(N, p) \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{|\int_0^1 \nabla v(x+t(y-x)) - \nabla v(x+(t-1)(y-x)) dt|^p |x-y|^p}{|x-y|^{2p}} \rho_n(x-y) dy \right) dx \\
&\leq M(N, p) \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{\int_0^1 |\nabla v(x+t(y-x)) - \nabla v(x+(t-1)(y-x))|^p dt}{|x-y|^p} \rho_n(x-y) dy \right) dx \\
&\leq M(N, p) \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \left( \int_0^1 \int_0^1 |\nabla^2 v(x+(t+s-1)(y-x))|^p ds dt \right) \rho_n(x-y) dy \right) dx \\
&= M(N, p) \int_0^1 \int_0^1 \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} (|\nabla^2 v(x+(t+s-1)(y-x))|^p) \rho_n(x-y) dy \right) dx ds dt \\
&= M(N, p) \int_0^1 \int_0^1 \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} (|\nabla^2 v(x+(t+s-1)\xi)|^p) \rho_n(\xi) d\xi \right) dx ds dt, \\
&= M(N, p) \int_0^1 \int_0^1 \int_{\mathbb{R}^N} \left( \rho_n(\xi) \int_{\mathbb{R}^N} (|\nabla^2 v(x+(t+s-1)\xi)|^p) dx \right) d\xi ds dt, \\
&= M(N, p) \int_0^1 \int_0^1 \int_{\mathbb{R}^N} \rho_n(\xi) \|\nabla^2 v\|_{L^p(\mathbb{R}^N)}^p d\xi ds dt \\
&= M(N, p) \|\nabla^2 v\|_{L^p(\mathbb{R}^N)}^p.
\end{aligned}$$

Consider now a sequence  $(v_k)_{k \in \mathbb{N}}$  in  $C^\infty(\mathbb{R}^N) \cap W^{2,p}(\mathbb{R}^N)$  approximating  $u$  in  $W^{2,p}(\mathbb{R}^N)$ . We already have from above that

$$(3.16) \quad \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{|d^2 v_k(x, y)|^p}{|x-y|^{2p}} \rho_n(x-y) dy \right) dx \leq M \|\nabla^2 v_k\|_{L^p(\mathbb{R}^N)}, \quad \forall k \in \mathbb{N},$$

or

$$(3.17) \quad \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{|v_k(x+h) - 2v_k(x) + v_k(x-h)|^p}{|h|^{2p}} \rho_n(h) dh \right) dx \leq M \|\nabla^2 v_k\|_{L^p(\mathbb{R}^N)}, \quad \forall k \in \mathbb{N}.$$

Since  $v_k$  converges to  $u$  in  $L^p(\mathbb{R}^N)$  we have that there exists a subsequence  $v_{k_\ell}$  converging to  $u$  almost everywhere. From an application of Fatou's lemma we get that for every  $x \in \mathbb{R}^N$

$$\underbrace{\int_{\mathbb{R}^N} \frac{|u(x+h) - 2u(x) + u(x-h)|^p}{|h|^{2p}} \rho_n(h) dh}_{F(x)} \leq M \liminf_{\ell \rightarrow \infty} \underbrace{\int_{\mathbb{R}^N} \frac{|v_{k_\ell}(x+h) - 2v_{k_\ell}(x) + v_{k_\ell}(x-h)|^p}{|h|^{2p}} \rho_n(h) dh}_{F_{k_\ell}(x)}.$$

Applying one more time Fatou's Lemma to  $F_{k_\ell}$ ,  $F$  we get that

$$\begin{aligned} & \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{|u(x+h) - 2u(x) + u(x-h)|^p}{|h|^{2p}} \rho_n(h) dh dx \\ & \leq M \liminf_{\ell \rightarrow \infty} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{|v_{k_\ell}(x+h) - 2v_{k_\ell}(x) + v_{k_\ell}(x-h)|^p}{|h|^{2p}} \rho_n(h) dh dx \\ & \leq M \liminf_{\ell \rightarrow \infty} \|\nabla^2 v_{k_\ell}\|_{L^p(\mathbb{R}^N)} \\ & = M \|\nabla^2 u\|_{L^p(\mathbb{R}^N)}. \end{aligned}$$

This argument, along with Jensen's inequality, allows us to conclude that the conditions of Theorem 1.3 are satisfied, in particular that  $H_n u$  is well-defined as a Lebesgue integral, so that the estimate (3.15) holds for  $W^{2,p}$  functions as well, thus completing the proof.

We now have the necessary tools to prove the localization for  $W^{2,p}$  functions.

*Proof of Theorem 1.5.* The result holds for functions  $v \in C_c^2(\mathbb{R}^N)$  since from Theorem 1.4 we have that  $H_n v \rightarrow \nabla^2 v$  in  $L^p(\mathbb{R}^N, \mathbb{R}^{N \times N})$ . We use now the fact that that  $C_c^\infty(\mathbb{R}^N)$  and hence  $C_c^2(\mathbb{R}^N)$  is dense in  $W^{2,p}(\mathbb{R}^N)$ , see for example [Bre83]. Let  $\epsilon > 0$ , then from density we have that there exists a function  $v \in C_c^2(\mathbb{R}^N)$  such that

$$\|\nabla^2 u - \nabla^2 v\|_{L^p(\mathbb{R}^N)} \leq \epsilon.$$

Thus using also Lemma 3.1 we have

$$\begin{aligned} \|H_n u - \nabla^2 u\|_{L^p(\mathbb{R}^N)} & \leq \|H_n u - H_n v\|_{L^p(\mathbb{R}^N)} + \|H_n v - \nabla^2 v\|_{L^p(\mathbb{R}^N)} + \|\nabla^2 v - \nabla^2 u\|_{L^p(\mathbb{R}^N)} \\ & \leq C\epsilon + \|H_n v - \nabla^2 v\|_{L^p(\mathbb{R}^N)} + \epsilon. \end{aligned}$$

Taking limits as  $n \rightarrow \infty$  we get

$$\limsup_{n \rightarrow \infty} \|H_n u - \nabla^2 u\|_{L^p(\mathbb{R}^N)} \leq (C+1)\epsilon,$$

and thus we conclude that

$$\lim_{n \rightarrow \infty} \|H_n u - \nabla^2 u\|_{L^p(\mathbb{R}^N)} = 0.$$

**3.3. Localization – BV<sup>2</sup>( $\mathbb{R}^N$ ) case.** Analogously with the first order case in [MS14], we can define a second order non-local divergence that corresponds to  $H_n$ , and we can also derive a second order non-local integration by parts formula which is an essential tool for the proofs of this section. The second order non-local divergence is defined for a function  $\phi = (\phi_{ij})_{i,j=1}^N$  as

$$\mathcal{D}_n^2 \phi(x) = \frac{N(N+2)}{2} \int_{\mathbb{R}^N} \frac{\phi(y) - 2\phi(x) + \phi(x+(x-y))}{|x-y|^2} \cdot \frac{\left((x-y) \otimes (x-y) - \frac{|x-y|^2}{N+2} I_N\right)}{|x-y|^2} \rho_n(x-y) dy. \quad (3.18)$$

where  $A \cdot B = \sum_{i,j=1}^N A_{ij} B_{ij}$  for two  $N \times N$  matrices  $A$  and  $B$ . Notice that (3.18) is well-defined for  $\phi \in C_c^2(\mathbb{R}^N, \mathbb{R}^{N \times N})$ .

**Theorem 3.3** (Second order non-local integration by parts formula). *Suppose that  $u \in L^1(\mathbb{R}^N)$ ,  $\frac{|d^2 u(x,y)|}{|x-y|^2} \rho_n(x-y) \in L^1(\mathbb{R}^N \times \mathbb{R}^N)$  and let  $\phi \in C_c^2(\mathbb{R}^N, \mathbb{R}^{N \times N})$ . Then*

$$(3.19) \quad \int_{\mathbb{R}^N} H_n u(x) \cdot \phi(x) dx = \int_{\mathbb{R}^N} u(x) \mathcal{D}_n^2 \phi(x) dx.$$

In fact, this theorem can be deduced as a consequence of Theorem 1.3 through a component by component application and collection of terms. The following lemma shows the convergence of the second order non-local divergence to the continuous analogue  $\text{div}^2 \phi$ , where  $\phi \in C_c^2(\mathbb{R}^N, \mathbb{R}^{N \times N})$  and

$$\text{div}^2 \phi := \sum_{i,j=1}^N \frac{\partial \phi_{ij}}{\partial x_i \partial x_j}.$$

**Lemma 3.4.** *Let  $\phi \in C_c^2(\mathbb{R}^N, \mathbb{R}^{N \times N})$ . Then for every  $1 \leq p \leq \infty$  we have*

$$(3.20) \quad \lim_{n \rightarrow \infty} \|\mathcal{D}_n^2 \phi - \operatorname{div}^2 \phi\|_{L^p(\mathbb{R}^N)} = 0.$$

*Proof.* The proof follows immediately from Theorem 1.4 and 1.5.  $\square$

Next we prove Lemma 3.2 that shows that the non-local Hessian (1.9) is well defined for  $u \in \operatorname{BV}^2(\mathbb{R}^N)$ . It is the analogue of Lemma 3.1 for functions in  $\operatorname{BV}^2(\mathbb{R}^N)$  this time.

*Proof of Lemma 3.2.* Let  $(u_k)_{k \in \mathbb{N}}$  be a sequence of functions in  $C^\infty(\mathbb{R}^N)$  that converges strictly in  $\operatorname{BV}^2(\mathbb{R}^N)$ . By the same calculations as in the proof of Lemma 3.1 we have for every  $k \in \mathbb{N}$

$$\int_{\mathbb{R}^N} |H_n u_k(x)| dx \leq M(N, 1) \|\nabla^2 u_k\|_{L^1(\mathbb{R}^N)}.$$

Using Fatou's Lemma in a similar way as in Lemma 3.1, we can establish that  $H_n u$  is well-defined as a Lebesgue integral, along with the estimate

$$\begin{aligned} \int_{\mathbb{R}^N} |H_n u(x)| dx &\leq M(N, 1) \liminf_{k \rightarrow \infty} \int_{\mathbb{R}^N} |D^2 u_k| \\ &= M(N, 1) \int_{\mathbb{R}^N} |D^2 u|. \end{aligned}$$

However, employing the strict convergence of  $D^2 u_k$  to  $D^2 u$ , the result is demonstrated.

We can now proceed in proving the localization result for  $\operatorname{BV}^2$  functions. We define  $\mu_n$  to be the  $\mathbb{R}^{N \times N}$ -valued finite Radon measures  $\mu_n := H_n u \mathcal{L}^N$ .

*Proof of Theorem 1.6.* Since  $C_c^\infty(\mathbb{R}^N, \mathbb{R}^{N \times N})$  is dense in  $C_0(\mathbb{R}^N, \mathbb{R}^{N \times N})$  it suffices to prove (1.12) for every  $\psi \in C_c^\infty(\mathbb{R}^N, \mathbb{R}^{N \times N})$ . Indeed, suppose we have done this, let  $\epsilon > 0$  and let  $\phi \in C_0(\mathbb{R}^N, \mathbb{R}^{N \times N})$  and  $\psi \in C_c^\infty(\mathbb{R}^N, \mathbb{R}^{N \times N})$  such that  $\|\phi - \psi\|_\infty < \epsilon$ . Then, using also the estimate (3.5), we have

$$\begin{aligned} \left| \int_{\mathbb{R}^N} H_n u(x) \cdot \phi(x) dx - \int_{\mathbb{R}^N} \phi(x) dD^2 u \right| &\leq \left| \int_{\mathbb{R}^N} H_n u(x) \cdot (\phi(x) - \psi(x)) dx \right| \\ &\quad + \left| \int_{\mathbb{R}^N} H_n u(x) \cdot \psi(x) dx - \int_{\mathbb{R}^N} \psi(x) dD^2 u \right| \\ &\quad + \left| \int_{\mathbb{R}^N} (\phi(x) - \psi(x)) dD^2 u \right| \\ &\leq \epsilon \int_{\mathbb{R}^N} |H_n u(x)| dx + \left| \int_{\mathbb{R}^N} H_n(x) \cdot \psi(x) dx - \int_{\mathbb{R}^N} \psi(x) dD^2 u \right| \\ &\quad + \epsilon \int_{\mathbb{R}^N} |D^2 u| \\ &\leq M\epsilon \int_{\mathbb{R}^N} |D^2 u| + \left| \int_{\mathbb{R}^N} H_n(x) \cdot \psi(x) dx - \int_{\mathbb{R}^N} \psi(x) dD^2 u \right| \\ &\quad + \epsilon \int_{\mathbb{R}^N} |D^2 u|. \end{aligned}$$

Taking the limit  $n \rightarrow \infty$  from both sides of the above inequality we get

$$\limsup_{n \rightarrow \infty} \left| \int_{\mathbb{R}^N} H_n u(x) \cdot \phi(x) dx - \int_{\mathbb{R}^N} \phi(x) dD^2 u \right| \leq \tilde{M}\epsilon$$

and since  $\epsilon$  is arbitrary we have (1.12). We thus proceed to prove (1.12) for  $C_c^\infty$  functions. From the estimate (3.5) we have that  $(|\mu_n|)_{n \in \mathbb{N}}$  is bounded, thus there exists a subsequence  $(\mu_{n_k})_{k \in \mathbb{N}}$  and a  $\mathbb{R}^{N \times N}$ -valued Radon measure  $\mu$  such that  $\mu_{n_k}$  converges to  $\mu$  weakly\*. This means that for every  $\psi \in C_c^\infty(\mathbb{R}^N, \mathbb{R}^{N \times N})$  we have

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}^N} H_{n_k} u(x) \cdot \psi(x) dx = \int_{\mathbb{R}^N} \psi(x) \cdot d\mu.$$

On the other hand from the integration by parts formula (3.19) and Lemma 3.4 we get

$$\begin{aligned} \lim_{k \rightarrow \infty} \int_{\mathbb{R}^N} H_{n_k} u(x) \cdot \psi(x) dx &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^N} u(x) \mathcal{D}_{n_k}^2 \psi(x) dx \\ &= \int_{\mathbb{R}^N} u(x) \operatorname{div}^2 \psi(x) dx \\ &= \int_{\mathbb{R}^N} \psi(x) \cdot dD^2 u. \end{aligned}$$

This means that  $\mu = D^2 u$ . Observe now that since we actually deduce that every subsequence of  $(\mu_n)_{n \in \mathbb{N}}$  has a further subsequence that converges to  $D^2 u$  weakly\*, then that the initial sequence  $(\mu_n)_{n \in \mathbb{N}}$  converges to  $D^2 u$  weakly star.

Let us note here that in the case  $N = 1$ , we can also prove strict convergence of the measures  $\mu_n$  to  $D^2 u$ , that is, in addition to (1.12) we also have

$$|\mu_n|(\mathbb{R}) \rightarrow |D^2 u|(\mathbb{R}).$$

**Theorem 3.5.** *Let  $N = 1$ . Then the sequence  $(\mu_n)_{n \in \mathbb{N}}$  converges to  $D^2 u$  strictly as measures, i.e.,*

$$(3.21) \quad \mu_n \rightarrow D^2 u, \quad \text{weakly}^*,$$

and

$$(3.22) \quad |\mu_n|(\mathbb{R}) \rightarrow |D^2 u|(\mathbb{R}).$$

*Proof.* The weak\* convergence was proven in Theorem 1.6. Since in the space of finite Radon measures, the total variation norm is lower semicontinuous with respect to the weak\* convergence, we also have

$$(3.23) \quad |D^2 u|(\mathbb{R}) \leq \liminf_{n \rightarrow \infty} |\mu_n|(\mathbb{R}).$$

Thus it suffices to show that

$$(3.24) \quad \limsup_{n \rightarrow \infty} |\mu_n|(\mathbb{R}) \leq |D^2 u|(\mathbb{R}).$$

Note that in dimension one the non-local Hessian formula is

$$(3.25) \quad H_n u(x) = \int_{\mathbb{R}} \frac{u(y) - 2u(x) + u(x + (x - y))}{|x - y|^2} \rho_n(x - y) dy.$$

Following the proof of Lemma 3.1, we can easily verify that for  $v \in C^\infty(\mathbb{R}) \cap \operatorname{BV}^2(\mathbb{R})$  we have

$$\int_{\mathbb{R}} |H_n v(x)| dx \leq \|\nabla^2 v\|_{L^1(\mathbb{R})},$$

i.e., the constant  $M$  that appears in the estimate (3.4) is equal to 1. Using Fatou's Lemma and the  $\operatorname{BV}^2$  strict approximation of  $u$  by smooth functions we get that

$$|\mu_n|(\mathbb{R}) = \int_{\mathbb{R}} |H_n u(x)| dx \leq |D^2 u|(\mathbb{R}),$$

from where (3.24) straightforwardly follows.  $\square$

**3.4. Characterization of higher-order Sobolev and BV spaces.** Characterization of Sobolev and BV spaces in terms of non-local, derivative-free energies has been done so far only in the first order case, see [BBM01, Pon04b, Men12, MS14]. Here we characterize the spaces  $W^{2,p}(\mathbb{R}^N)$  and  $\operatorname{BV}^2(\mathbb{R}^N)$  using our definition of non-local Hessian.

*Proof of Theorem 1.7.* Firstly, we prove (1.13). Suppose that  $u \in W^{2,p}(\mathbb{R}^N)$ . Then, Lemma 3.1 gives

$$\liminf_{n \rightarrow \infty} \int_{\mathbb{R}^N} |H_n u(x)|^p dx \leq M \|\nabla^2 u\|_{L^p(\mathbb{R}^N)}^p < \infty.$$

Suppose now conversely that

$$\liminf_{n \rightarrow \infty} \int_{\mathbb{R}^N} |\mathfrak{H}_n u(x)|^p dx < \infty.$$

This means that up to a subsequence,  $\mathfrak{H}_n u$  is bounded in  $L^p(\mathbb{R}^N, \mathbb{R}^{N \times N})$ , thus there exists a subsequence  $(\mathfrak{H}_{n_k} u)_{k \in \mathbb{N}}$  and  $v \in L^p(\mathbb{R}^N, \mathbb{R}^{N \times N})$  such that  $\mathfrak{H}_{n_k} u \rightharpoonup v$  weakly in  $L^p(\mathbb{R}^N, \mathbb{R}^{N \times N})$ . Thus, using the definition of  $L^p$  weak convergence together with the definition of  $\mathfrak{H}_n u$ , we have for every  $\psi \in C_c^\infty(\mathbb{R}^N)$ ,

$$\begin{aligned} \int_{\mathbb{R}^N} v^{ij}(x) \cdot \psi(x) dx &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^N} \mathfrak{H}_{n_k}^{ij} u(x) \cdot \psi(x) \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^N} u(x) H_n^{ij} \psi(x) \\ &= \int_{\mathbb{R}^N} u(x) \frac{\partial^2 \psi(x)}{\partial x_i \partial x_j} dx, \end{aligned}$$

something that shows that  $v = \nabla^2 u$  is the second order weak derivative of  $u$ . Now since  $u \in L^p(\mathbb{R}^N)$  and the second order distributional derivative is a function, mollification of  $u$  and the Gagliardo-Nirenberg inequality (see [Nir59, p. 128, Equation 2.5])

$$(3.26) \quad \|\nabla u\|_{L^p(\mathbb{R}^N, \mathbb{R}^N)} \leq C \|\nabla^2 u\|_{L^p(\mathbb{R}^N, \mathbb{R}^{N \times N})}^{\frac{1}{2}} \|u\|_{L^p(\mathbb{R}^N)}^{\frac{1}{2}},$$

implies that the first distributional derivative and it belongs to  $L^p(\mathbb{R}^N, \mathbb{R}^N)$  and thus  $u \in W^{2,p}(\mathbb{R}^N)$ .

We now proceed in proving (1.14). Again supposing that  $u \in \text{BV}^2(\mathbb{R}^N)$  we have that Lemma 3.2 gives us

$$\liminf_{n \rightarrow \infty} \int_{\mathbb{R}^N} |H_n u(x)| dx \leq C |D^2 u|(\mathbb{R}^N).$$

Suppose now that

$$\liminf_{n \rightarrow \infty} \int_{\mathbb{R}^N} |\mathfrak{H}_n u(x)| dx < \infty.$$

Considering again the measures  $\mu_n = \mathfrak{H}_n u \mathcal{L}^N$  we have that there exists a subsequence  $(\mu_{n_k})_{k \in \mathbb{N}}$  and a finite Radon measure  $\mu$  such that  $\mu_{n_k} \xrightarrow{*} \mu$  weakly\*. Then for every  $\psi \in C_c^\infty(\mathbb{R}^N)$  we have, similarly as before,

$$\begin{aligned} \int_{\mathbb{R}^N} \psi d\mu^{ij} &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^N} \mathfrak{H}_{n_k}^{ij} u(x) \cdot \psi(x) dx \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^N} u(x) H_n^{ij} \psi(x) dx \\ &= \int_{\mathbb{R}^N} u(x) \frac{\partial^2 \psi(x)}{\partial x_i \partial x_j} dx. \end{aligned}$$

something that shows that  $\mu = D^2 u$ . Again, by first mollifying and then passing the limit, the inequality (3.26) implies that  $Du \in \mathcal{M}(\mathbb{R}^N, \mathbb{R}^N)$ . However,  $Du \in \mathcal{M}(\mathbb{R}^N, \mathbb{R}^N)$  and  $D^2 u \in \mathcal{M}(\mathbb{R}^N, \mathbb{R}^{N \times N})$  implies that  $Du$  is an  $L^1(\mathbb{R}^N, \mathbb{R}^N)$  function (see [AFP00, Exercise 3.2]), and we therefore conclude that  $u \in \text{BV}^2(\mathbb{R}^N)$ .

**3.5. Gamma convergence.** For notational convenience we define the functional

$$(3.27) \quad F_n(u) := \int_{\mathbb{R}^N} |\mathfrak{H}_n u| dx.$$

*Proof of Theorem 1.8.* The computation of the Gamma limit consists of two inequalities. For the lower bound, we must show that

$$|D^2 u|(\mathbb{R}^N) \leq \liminf_{n \rightarrow \infty} F_n(u_n)$$

for every sequence  $u_n \rightarrow u$  in  $L^1(\mathbb{R}^N)$ . Without loss of generality we may assume that

$$C := \liminf_{n \rightarrow \infty} F_n(u_n) < +\infty.$$

which implies that

$$\sup_{\varphi} \liminf_{n \rightarrow \infty} \left| \int_{\mathbb{R}^N} \mathfrak{H}_n^{ij} u_n \varphi dx \right| \leq C.$$

where the supremum is taken over  $\varphi \in C_c^\infty(\mathbb{R}^N)$  such that  $\|\varphi\|_{L^\infty(\mathbb{R}^N)} \leq 1$ . Now, the definition of the distributional non-local Hessian and the convergence  $u_n \rightarrow u$  in  $L^1(\mathbb{R}^N)$  imply that

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{\mathbb{R}^N} \mathfrak{H}_n^{ij} u_n \varphi \, dx &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}^N} u_n H_n^{ij} \varphi \, dx \\ &= \int_{\mathbb{R}^N} u \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \, dx. \end{aligned}$$

We thus conclude that

$$\sup_{\varphi} \liminf_{n \rightarrow \infty} \left| \int_{\mathbb{R}^N} u \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \, dx \right| \leq C,$$

which, arguing as in the previous section, says that  $u \in \text{BV}^2(\mathbb{R}^N)$ , in particular that  $D^2u \in \mathcal{M}(\mathbb{R}^N, \mathbb{R}^{N \times N})$  and

$$|D^2u|(\mathbb{R}^N) \leq \Gamma_{L^1(\mathbb{R}^N)}^- \liminf_{n \rightarrow \infty} F_n(u)$$

for every  $u \in L^1(\mathbb{R}^N)$ .

For the upper bound we observe that if  $u \in C_c^2(\mathbb{R}^N)$ , we have by the uniform convergence of Theorem 1.4 and the fact that  $u$  is sufficiently smooth with compact support that

$$\lim_{n \rightarrow \infty} F_n(u) = |D^2u|(\mathbb{R}^N).$$

Then choosing  $u_n = u$  we conclude

$$\begin{aligned} \Gamma_{L^1(\mathbb{R}^N)}^- \limsup_{n \rightarrow \infty} F_n(u) &\leq \lim_{n \rightarrow \infty} F_n(u) \\ &= |D^2u|(\mathbb{R}^N). \end{aligned}$$

Now, taking the lower semicontinuous envelope with respect to  $L^1(\mathbb{R}^N)$  strong convergence, using  $\Gamma_{L^1(\mathbb{R}^N)}^- \limsup$  is lower semicontinuous on  $L^1(\mathbb{R}^N)$  (c.f. [DM93, Proposition 6.8]), we deduce that

$$\begin{aligned} \Gamma_{L^1(\Omega)}^- \limsup_{n \rightarrow \infty} F_n(u) &\leq sc_{L^1(\Omega)}^- |D^2u|(\mathbb{R}^N). \\ &= |D^2u|(\mathbb{R}^N) \end{aligned}$$

for all  $u \in L^1(\mathbb{R}^N)$ .

#### 4. EXTENSIONS AND APPLICATIONS

**4.1. An asymmetric extension.** In the previous sections we have shown that our non-local definition of  $H_n$  as in (1.9) localizes to the classical distributional Hessian for a specific choice of the weights  $\rho_n$ , and thus can be rightfully called a non-local Hessian. In numerical applications, however, the strength of such non-local models lies in the fact that the weights can be chosen to have non-local interactions and model specific patterns in the data. A classic example is the non-local total variation [?]:

$$(4.1) \quad J_{NL-TV}(u) = \int_{\Omega} \int_{\Omega} |u(x) - u(y)| \sqrt{w(x,y)} \, dy \, dx.$$

A possible choice is to choose  $w(x,y)$  large if the *patches* (neighbourhoods) around  $x$  and  $y$  are similar with respect to a *patch distance*  $d_a$ , such as a weighted  $\ell^2$  norm, and small if they are not. In [?] this is achieved by setting  $w(x,y) = 1$  if the neighbourhood around  $y$  is one of the  $K \in \mathbb{N}$  closest to the neighbourhood around  $x$  in a search window, and  $w(x,y) = 0$  otherwise. In effect, if the image contains a repeating pattern with a defect that is small enough to not throw off the patch distances too much, it will be repaired as long as most similar patterns do not show the defect.

Computing suitable weights is much less obvious in the case of  $H$ . We can formally extend (1.9) using arbitrary pairwise weights  $\rho : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$(4.2) \quad H_{\rho}u(x) = C \int_{\mathbb{R}^N} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \frac{\left(zz^{\top} - \frac{|z|^2}{N+2}I_N\right)}{|z|^2} \rho_x(z) \, dz,$$

and use it to create non-local generalisations of functionals such as  $\text{TV}^2$ , for example to minimise the non-local  $L^2 - \text{TV}^2$  model

$$(4.3) \quad f(u) := \int_{\mathbb{R}^N} |u - g|^2 dx + \alpha \int_{\mathbb{R}^N} |\mathcal{H}_\rho| dx.$$

However, apart from being formulated on  $\mathbb{R}^N$  instead of  $\Omega$ , formulation (4.2) has an important drawback compared to the first-order formulation (4.1): while the weights are defined between two points  $x$  and  $y$ , the left part of the integrand uses the values of  $u$  not only at  $x$  and  $y$ , but also at the “mirrored” point  $x + (x - y)$ . In fact we can replace the weighting function by the symmetrised version  $\frac{1}{2}\{\rho_x(y - x) + \rho_x(x - y)\}$ , which in effect relates three points instead of two, and limits the choice of possible weighting functions.

In this section we therefore introduce a more versatile extension of (1.9) that allows for full non-symmetric weights. We start with the realisation that the finite-differences integrand in (4.2) effectively comes from canceling the first-order differences in the Taylor expansion of  $u$  around  $x$ , which couples the values of  $u$  at  $x$ ,  $y$ , and  $x + (x - y)$  into one term. Instead, we can avoid this coupling by directly defining the non-local gradient  $G'_u(x)$  and Hessian looking for a non-local gradient  $G_u(x) \in \mathbb{R}^N$  and Hessian  $H_u(x) \in \text{Sym}(\mathbb{R}^{N \times N})$  that *best explain  $u$  around  $x$  in terms of a quadratic model*, i.e., that take the place of the gradient and Hessian in the Taylor expansion:

$$(4.4) \quad (G'_u(x), H'_u(x)) := \underset{G_u \in \mathbb{R}^N, H_u \in \text{Sym}(\mathbb{R}^{N \times N})}{\text{argmin}} \frac{1}{2} \int_{\Omega - \{x\}} \left( u(x+z) - u(x) - G_u^\top z - \frac{1}{2} z^\top H_u z \right)^2 \sigma_x(z) dz.$$

Here the variable  $x + z$  takes the place of  $y$  in (1.9). We denote definition (4.4) the *implicit* non-local Hessian as opposed to the *explicit* formulation (4.2).

The advantage is that any terms involving  $\sigma_x(z)$  are now only based on the two values of  $u$  at  $x$  and  $y = x + z$ , and (in particular bounded) domains other than  $\mathbb{R}^N$  are naturally dealt with, which is important for a numerical implementation. We also note that this approach allows to incorporate non-local first-order terms as a side-effect, and can be naturally extended to third- and higher-order derivatives, which we leave to further work.

With respect to implementation, the implicit model (4.4) does not add much to the overall difficulty: it is enough to add the non-local gradient and Hessian  $G'_u(x) \in \mathbb{R}^N$  and  $H'_u(x) \in \mathbb{R}^{N \times N}$  as additional variables to the problem, and couple them to  $u$  by adding the optimality conditions for (4.4) to the problem. Since (4.4) is a finite-dimensional quadratic minimization problem, the optimality conditions are linear, i.e., of the form  $A_{u,x} H'_u(x) = b_{u,x}$ . Such linear constraints can be readily added to most solvers capable of solving the local problem. Alternatively the matrices  $A_{u,x}$  can be inverted explicitly in a pre-computation step, however we did not find this to increase overall performance.

While the implicit and explicit model look different at first glance, from the considerations about the Taylor expansion we expect them to be closely related, and in fact this is true as shown in the following series of propositions. The first proposition explicitly computes the constants  $C_{ij}$  and is needed for the proofs:

**Proposition 4.1.** *For  $N \geq 1$  we have the following definition and identity for  $C$ :*

$$C_{i_0 j_0} := \int_{S^{N-1}} \nu_{i_0}^2 \nu_{j_0}^2 d\mathcal{H}^{N-1}(x) = \frac{|S^{N-1}|}{N(N+2)} \cdot \begin{cases} 1, & i_0 \neq j_0, \\ 3, & i_0 = j_0. \end{cases}$$

The matrix  $C = (C_{i_0 j_0})_{i_0, j_0=1, \dots, N}$  is

$$C = C_{12}(E + 2I),$$

where  $E \in \mathbb{R}^{N \times N}$  is the all-ones matrix.

*Proof.* See Appendix. □

The following theorem shows that for radially symmetric  $\rho_x$  and  $\sigma_x$ , if we restrict ourselves to a circle, the *implicit* non-local Hessian is (up to a factor) the *explicit* non-local Hessian as defined earlier.

**Theorem 4.2.** *Let  $\Omega := \mathbb{R}^N$ ,  $h > 0$ ,  $x \in \mathbb{R}^N$ . Consider the minimization problem*

$$(4.5) \quad \min_{G_{u,h} \in \mathbb{R}^N, H_{u,h} \in \text{Sym}(\mathbb{R}^{N \times N})} \frac{1}{2} \int_{hS^{N-1}} \left( u(x+z) - u(x) - G_{u,h}^\top z - \frac{1}{2} z^\top H_{u,h} z \right)^2 d\mathcal{H}^{N-1}(z),$$

and assume that the integral is finite for every  $G_{u,h} \in \mathbb{R}^N$  and  $H_{u,h} \in \mathbb{R}^{N \times N}$ . Then the non-local Hessian part of the minimizing pair  $(G'_{u,h}, H'_{u,h})$  is given by

$$(4.6) \quad H'_{u,h} = \frac{1}{2} C_{12}^{-1} h^{-(N-1)} \int_{h\mathcal{S}^{N-1}} \left( \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \right) \left( \frac{zz^\top}{|z|^2} - \frac{1}{N+2} I \right) d\mathcal{H}^{N-1}(z).$$

*Proof.* The optimality conditions for (4.5) in terms of  $G_u = G_{u,h}$  and  $H_u = H_{u,h}$ , leaving the assumption on the symmetry of  $H_u$  aside for the moment as we will enforce them explicitly later, are

$$\begin{aligned} 0 &= \int_{h\mathcal{S}^{N-1}} - \left( u(x+z) - u(x) - G_u^\top z - \frac{1}{2} z^\top H_u z \right) z d\mathcal{H}^{N-1}(z), \\ 0 &= \int_{h\mathcal{S}^{N-1}} - \frac{1}{2} \left( u(x+z) - u(x) - G_u^\top z - \frac{1}{2} z^\top H_u z \right) z z d\mathcal{H}^{N-1}(z), \end{aligned}$$

Thus from the first line we get  $N$  equations

$$\begin{aligned} &\left\langle \int_{h\mathcal{S}^{N-1}} z_i z d\mathcal{H}^{N-1}(z), G_u \right\rangle + \left\langle \int_{h\mathcal{S}^{N-1}} z_i \left( \frac{1}{2} z z^\top \right) d\mathcal{H}^{N-1}(z), H_u \right\rangle \\ &= \int_{h\mathcal{S}^{N-1}} z_i (u(x+z) - u(x)) d\mathcal{H}^{N-1}(z), \quad i = 1, \dots, N. \end{aligned}$$

Similarly, from the second line we get  $N^2$  equations (of which some may be redundant)

$$\begin{aligned} &\left\langle \int_{h\mathcal{S}^{N-1}} \frac{1}{2} z_i z_j z^\top d\mathcal{H}^{N-1}(z), G_u \right\rangle + \left\langle \int_{h\mathcal{S}^{N-1}} \frac{1}{2} z_i z_j \left( \frac{1}{2} z z^\top \right) d\mathcal{H}^{N-1}(z), H_u \right\rangle \\ &= \int_{h\mathcal{S}^{N-1}} \frac{1}{2} z_i z_j (u(x+z) - u(x)) d\mathcal{H}^{N-1}(z), \quad i, j = 1, \dots, N. \end{aligned}$$

Note that  $G_u$  and  $H_u$  are elements of  $\mathbb{R}^N$  and  $\mathbb{R}^{N \times N}$ , and both sides of the inner products are finite-dimensional vectors respective matrices.

If we collect the entries of  $G_u$  and  $H_u$  in a vector  $p = (p_G, p_H)$ , these two sets of equations can be rewritten as a linear system with an  $m \times m$  block matrix,

$$\begin{pmatrix} A & V^\top \\ V & B \end{pmatrix} p = \begin{pmatrix} a \\ b \end{pmatrix}.$$

The entries in the sub-matrices  $V$  are all of the form

$$\int_{h\mathcal{S}^{N-1}} z_i (z_j z_k) d\mathcal{H}^{N-1}(z), \quad i, j, k \in \{1, \dots, n\}.$$

No matter what the choice of  $i, j, k$  is, there is always at least one index with an odd power, so every single one of these integrals is zero due to symmetry. This means that the conditions on the gradient and the Hessian parts of  $p$  decouple, i.e., the problem is

$$(4.7) \quad Ap_G = a, \quad Bp_H = b.$$

We can therefore look at the isolated problem of computing the Hessian part  $p_H$ , or equivalently  $H_u$ , without interference from the gradient part. The matrix  $B$  is of the form

$$\int_{h\mathcal{S}^{N-1}} \frac{1}{2} z_i z_j \frac{1}{2} z_{i_0} z_{j_0} d\mathcal{H}^{N-1}(z).$$

Again due to symmetry, the only way that this integral is non-zero is if there are no odd powers, so either all indices are the same or there are two pairs:

$$\begin{aligned} &\frac{1}{4} \int_{h\mathcal{S}^{N-1}} z_k^4 d\mathcal{H}^{N-1}(z), \quad \text{for some } k, \\ \text{or} &\frac{1}{4} \int_{h\mathcal{S}^{N-1}} z_k^2 z_l^2 d\mathcal{H}^{N-1}(z), \quad \text{for some } k \neq l. \end{aligned}$$

The vector  $b$  in (4.7) is

$$\int_{h\mathcal{S}^{N-1}} (u(x+z) - u(x)) \frac{1}{2} z_i z_j d\mathcal{H}^{N-1}(z) = \frac{1}{2} \int_{h\mathcal{S}^{N-1}} (u(x+z) - 2u(x) + u(x-z)) \frac{1}{2} z_i z_j d\mathcal{H}^{N-1}(z).$$

Now the optimality conditions on the  $p_H = (p_{ij})_{i,j=1,\dots,N}$  part of  $p$  are of the form (for all  $i_0, j_0$ )

$$(4.8) \quad \begin{aligned} & \sum_{i,j} p_{ij} \int_{h\mathcal{S}^{N-1}} \frac{1}{2} z_i z_j \frac{1}{2} z_{i_0} z_{j_0} d\mathcal{H}^{N-1}(z) \\ &= \frac{1}{2} \int_{h\mathcal{S}^{N-1}} (u(x+z) - 2u(x) + u(x-z)) \frac{1}{2} z_{i_0} z_{j_0} d\mathcal{H}^{N-1}(z), \quad i_0, j_0 = 1, \dots, N, \end{aligned}$$

or alternatively

$$\begin{aligned} & \sum_{i,j} p_{ij} \int_{\mathcal{S}^{N-1}} \frac{1}{2} h\nu_i h\nu_j \frac{1}{2} h\nu_{i_0} h\nu_{j_0} d\mathcal{H}^{N-1}(\nu) \\ &= \frac{1}{4} \int_{\mathcal{S}^{N-1}} (u(x+h\nu) - 2u(x) + u(x-h\nu)) h\nu_{i_0} h\nu_{j_0} d\mathcal{H}^{N-1}(\nu), \quad i_0, j_0 = 1, \dots, N. \end{aligned}$$

We can divide everything by the constant  $\frac{1}{4}h^4$  to get

$$(4.9) \quad \begin{aligned} & \sum_{i,j} p_{ij} \int_{\mathcal{S}^{N-1}} \nu_i \nu_j \nu_{i_0} \nu_{j_0} d\mathcal{H}^{N-1}(\nu) \\ &= \int_{\mathcal{S}^{N-1}} \left( \frac{u(x+h\nu) - 2u(x) + u(x-h\nu)}{h^2} \right) \nu_{i_0} \nu_{j_0} d\mathcal{H}^{N-1}(\nu), \quad i_0, j_0 = 1, \dots, N. \end{aligned}$$

We consider the different cases for  $i_0, j_0$  separately.

- If  $i_0 \neq j_0$ , then on the left-hand side all terms vanish except for  $(i, j) = (i_0, j_0)$  and  $(i, j) = (j_0, i_0)$ . Overall we get the condition

$$(p_{i_0 j_0} + p_{j_0 i_0}) \int_{\mathcal{S}^{N-1}} \nu_{i_0}^2 \nu_{j_0}^2 d\mathcal{H}^{N-1}(\nu) = \int_{\mathcal{S}^{N-1}} \left( \frac{u(x+h\nu) - 2u(x) + u(x-h\nu)}{h^2} \right) \nu_{i_0} \nu_{j_0} d\mathcal{H}^{N-1}(\nu).$$

Since we require  $H_u = p_{ij}$  to be symmetric, we get an explicit solution for  $p_{i_0 j_0}$ :

$$p_{i_0 j_0} = \frac{1}{2} C_{i_0 j_0}^{-1} \int_{\mathcal{S}^{N-1}} \left( \frac{u(x+h\nu) - 2u(x) + u(x-h\nu)}{h^2} \right) \nu_{i_0} \nu_{j_0} d\mathcal{H}^{N-1}(\nu),$$

where  $C$  is the scalar as in Prop. 4.1:

$$C_{i_0 j_0} := \int_{\mathcal{S}^{N-1}} \nu_{i_0}^2 \nu_{j_0}^2 d\mathcal{H}^{N-1}(\nu).$$

- If  $i_0 = j_0$ , then on the left-hand side all terms vanish except for  $i = j$  and we get

$$\sum_i p_{ii} \int_{\mathcal{S}^{N-1}} \nu_i^2 \nu_{i_0}^2 d\mathcal{H}^{N-1}(\nu) = \int_{\mathcal{S}^{N-1}} \left( \frac{u(x+h\nu) - 2u(x) + u(x-h\nu)}{h^2} \right) \nu_{i_0} \nu_{j_0} d\mathcal{H}^{N-1}(\nu).$$

With the definition for  $C$  we can simplify the left-hand side to

$$\sum_i p_{ii} C_{ii} = \int_{\mathcal{S}^{N-1}} \left( \frac{u(x+h\nu) - 2u(x) + u(x-h\nu)}{h^2} \right) \nu_{i_0} \nu_{i_0} d\mathcal{H}^{N-1}(\nu), \quad i_0 = 1, \dots, n.$$

Consequently, for the diagonal vector  $p' = (p_{11}, \dots, p_{NN})$  and  $y = (\text{rhs})_{i_0}$  we get from Proposition 4.1

$$C_{12}(E + 2I)p' = y$$

with  $E$  the all-ones matrix in  $\mathbb{R}^{N \times N}$ . We have  $(E + 2I)^{-1} = \frac{1}{2} \left( I - \frac{1}{N+2} E \right)$ , thus we can rewrite

$$C_{12}p' = \frac{1}{2} \left( I - \frac{1}{N+2} E \right) y.$$

The  $E$  term in the right-hand side is (since  $|\nu|^2 = 1$  on  $\mathcal{S}^{N-1}$ )

$$\begin{aligned} & -\frac{1}{N+2} \sum_i \int_{\mathcal{S}^{N-1}} \left( \frac{u(x+h\nu) - 2u(x) + u(x-h\nu)}{h^2} \right) \nu_i^2 d\mathcal{H}^{N-1}(\nu) \\ &= -\frac{1}{N+2} \int_{\mathcal{S}^{N-1}} \left( \frac{u(x+h\nu) - 2u(x) + u(x-h\nu)}{h^2} \right) 1 d\mathcal{H}^{N-1}(\nu). \end{aligned}$$

Putting both cases together, we can now solve the least-squares problem for a fixed radius  $h$  in closed form for  $p_H$  and therefore the first part of the claimed identity:

$$(4.10) \quad p_H(h) = p = \frac{1}{2} C_{12}^{-1} \int_{\mathcal{S}^{N-1}} \left( \frac{u(x+h\nu) - 2u(x) + u(x-h\nu)}{h^2} \right) \left( \nu\nu^\top - \frac{1}{N+2} I \right) d\mathcal{H}^{N-1}(\nu)$$

The last part follows with the substitution  $z = h\nu$ .  $\square$

The following theorem shows that the explicit form  $H_u$  is the same as *separately* solving the least-squares problem on all spheres for all  $h > 0$  and then taking a weighted mean based on  $\rho$ .

**Proposition 4.3.** *Assume  $\rho_x$  is radially symmetric, i.e.,  $\rho_x(z) = \rho_x(|z|)$ . Then the explicit non-local Hessian  $H_u$  can be written as*

$$(4.11) \quad H_u = \int_0^\infty H'_{u,h} \int_{h\mathcal{S}^{N-1}} \rho_x(z) d\mathcal{H}^{N-1}(z) dh.$$

*Proof.* This is a direct consequence of Theorem. 4.2:

$$\begin{aligned} H_u &= \frac{N(N+2)}{2} \int_{\mathbb{R}^N} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \left( \frac{zz^\top}{|z|^2} - \frac{1}{N+2} I_N \right) \rho_x(z) dz \\ &= \frac{N(N+2)}{2} \int_0^\infty \rho_x(h) \int_{h\mathcal{S}^{N-1}} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \left( \frac{zz^\top}{|z|^2} - \frac{1}{N+2} I_N \right) dz dh \\ &\stackrel{\text{Thm. 4.2}}{=} \frac{N(N+2)}{2} \int_0^\infty \rho_x(h) 2C_{12} h^{N-1} H_{u,h} dh \\ &\stackrel{\text{Prop. 4.1}}{=} N(N+2) \frac{2\pi^{N/2}}{N(N+2)\Gamma(N/2)} \int_0^\infty \rho_x(h) h^{N-1} H_{u,h} dh \\ &= \frac{2\pi^{N/2}}{\Gamma(N/2)} \int_0^\infty h^{N-1} |h\mathcal{S}^{N-1}|^{-1} H_{u,h} \int_{h\mathcal{S}^{N-1}} \rho_x(z) dh \\ &= \frac{2\pi^{N/2}}{\Gamma(N/2)} |\mathcal{S}^{N-1}|^{-1} \int_0^\infty H_{u,h} \int_{h\mathcal{S}^{N-1}} \rho_x(z) dz dh \\ &= \int_0^\infty H_{u,h} \int_{h\mathcal{S}^{N-1}} \rho_x(z) dz dh. \end{aligned}$$

$\square$

Note that while Proposition 4.3 requires radial symmetry of  $\rho$ , this symmetry was generally assumed throughout Section 3. Therefore Section 3 can also be seen as providing localization results for *implicit* models of the form (4.11).

**4.2. Choosing the weights for jump preservation.** A characteristic of non-local models is that they are extremely flexible due to the many degrees of freedom in choosing the weights. In this work we will focus on improving on the question of how to reconstruct images that are piecewise quadratic but may have jumps. The issue here is that one wants to keep the Hessian sparse in order to favor piecewise affine functions, but doing it in a straightforward way, such as by adding  $|D^2u|(\Omega)$  as a regularizer, enforces too much first-order regularity [Sch98, LLT03, LT06, HS06].

There have been several attempts of overcoming this issue, most notably approaches based on combined first- and higher-order functionals (see [PS14] and the references therein), infimal convolution [CL97], and Total Generalized Variation [BKP10, SST11]. Here we propose another strategy making use of the non-local formulation (4.4).

We draw our motivation for choosing the weights partly from a recent discussion of non-local ‘‘Amoeba’’ filters [LDM07, WBV11, Wel12]. Amoeba filters use classical techniques such as iterated median filtering, but the structuring element is chosen in a highly adaptive local way that can follow the image structures, instead of being restricted to a small parametrized set of shapes. In the following we propose to extend this idea to the higher-order energy minimization framework (Figure 1).

Given a noisy image  $g : \Omega \rightarrow \mathbb{R}$ , we first compute its (equally noisy) gradient image  $\nabla g$  (in all of this section we consider only the discretized problem, so we can assume that the gradient exists). We then define

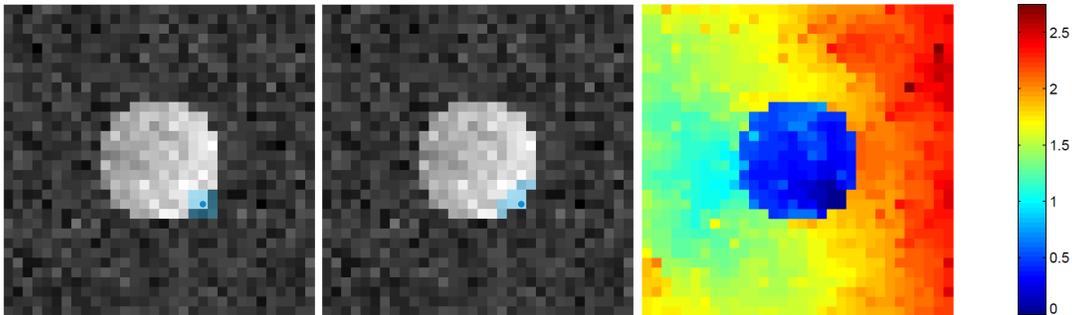


FIGURE 1. Adaptive choice of the neighbourhood for the image of a disc with constant slope and added Gaussian noise. **Left:** A standard discretization of the second-order derivatives at the marked point uses all points in a  $3 \times 3$  neighbourhood. **Center:** With a suitable choice of the weights  $\sigma_x$ , the discretization of the (non-local) Hessian at the marked point only (or mostly) involved points that are likely to belong to the same affine region, here the inside of the disc. This allows to use a straight-forward second-order regularizer such as the norm of the Hessian  $\|H'_u\|$ , while preserving jumps. **Right:** Geodesic distance  $d_{\mathcal{M}}$  from the marked point  $x$  in the lower right. Points that are separated from  $x$  by a strong edge have a large geodesic distance to  $x$ , and are therefore not included in the neighbourhood of  $x$  that is used to define the non-local Hessian at  $x$ .

the Riemannian manifold  $\mathcal{M}$  on the points of  $\Omega$  with the usual Riemannian metric, weighted by  $\varphi(\nabla g)$  for some function  $\varphi : \mathbb{R}^N \rightarrow \mathbb{R}_+$ . In our experiments we used

$$(4.12) \quad \varphi(g) := |g|^2 + \gamma$$

for small  $\gamma > 0$ , but other choices are equally possible. The choice of  $\gamma$  controls how strongly the edge information is taken into account: for  $\gamma = 0$  the weights are fully anisotropic, while for large  $\gamma$  the magnitude of the gradient becomes irrelevant, and the method will reduce to an isotropic regularization.

With these definitions, the *geodesic distance*  $d_{\mathcal{M}}$  between two points  $x, y \in \Omega$  now has a powerful interpretation: If  $d_{\mathcal{M}}(x, y)$  is large, this indicates that  $x$  and  $y$  are separated by a strong edge and should therefore not appear in the same regularization term. On the other hand, small  $d_{\mathcal{M}}(x)$  indicates that  $x$  and  $y$  are part of a more homogeneous region, and it is reasonable to assume that they should be part of the same affine part of the reconstructed image.

For a given point  $x \in \Omega$ , we sort the neighbours  $y^1, y^2, \dots$  in order of ascending distance, i.e.,  $d_{\mathcal{M}}(x, y^1) \leq d_{\mathcal{M}}(x, y^2) \leq \dots$ . We choose a neighbourhood size  $M \in \mathbb{N}$  and set the weights to

$$(4.13) \quad \sigma_x(y) := \begin{cases} 1, & i \leq M, \\ 0 & i > M. \end{cases}$$

In other words, the non-local Hessian at  $x$  is computed using its  $M$  closest neighbours with respect to the geodesic distance through the gradient image.

The geodesic distances  $\sigma_x(y)$  can be efficiently computed using the Fast Marching Method [Set99, OF03] by solving the Eikonal equation

$$(4.14) \quad |\nabla c(y)| = \varphi(\nabla g(y)),$$

$$(4.15) \quad c(x) = 0,$$

and setting  $d_{\mathcal{M}}(x, y) = c(x)$ . Although it is necessary to process this step for every point  $x \in \Omega$ , it is in practice a relatively cheap operation: the Fast Marching Method visits the neighbours of  $x$  in the order of ascending distance  $d_{\mathcal{M}}$ , which means it can be stopped after  $M$  points, with  $M$  usually between 5 and 20. If  $M$  is chosen too small, one risks that the linear equation system that defines the non-local Hessian in (4.4) becomes underdetermined. In our experiments we found  $M = 12$  to be a good compromise, but the choice

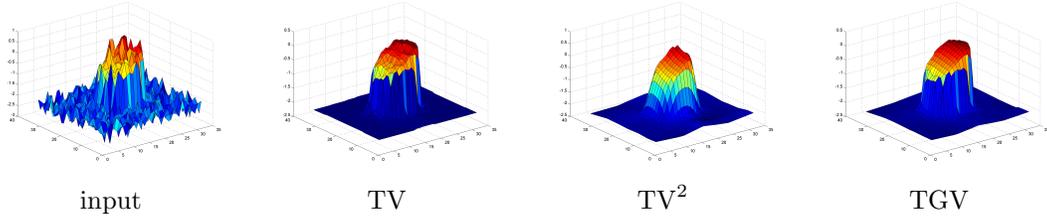


FIGURE 2. Classical local regularization. The input consists of a disc-shaped slope with additive Gaussian noise,  $\sigma = 0.25$ . Shown is the result of denoising the input with  $L^1$ -data term. Total variation (TV,  $\alpha = 1.1$ ) regularization generates the well-known staircasing effect. Regularization of the Hessian (TV<sup>2</sup>,  $\alpha = 0.8$ ) avoids this problem, at the cost of over-smoothing jumps. Total Generalized Variation (TGV,  $\alpha_0 = 1.5, \alpha_1 = 1.1$ ) performs best, but still clips the slope at the top.

does not appear to be a very critical one. In the experiments we used the  $L^1$ -non-local TV<sup>2</sup> model

$$(4.16) \quad \min_{u: \Omega \rightarrow \mathbb{R}} \sum_{x \in \Omega} |u(x) - g(x)|^p dx + \alpha \sum_{x \in \Omega} \omega(x) |H'_u(x)| dx$$

$$(4.17) \quad \text{s.t.} \quad AH'_u = Bu,$$

where  $\alpha > 0$  is the regularization strength, and  $p \in \{1, 2\}$ . The linear constraints implement the optimality conditions for  $H'_u$ .

The local weight  $\omega$  is set as  $\omega(x) = M/|\{y \in \Omega | B_{y,x} \neq 0\}|$ . While the approach does work with uniform weights  $\omega = 1$ , we found that in some cases it can erroneously leave single outlier points intact. We believe that this is caused by a subtle issue: by construction of  $\sigma$ , outlier points are usually close neighbour to fewer points. Therefore they appear in fewer of the regularization terms  $|H'_u(x)|$ , which effectively decreases regularization strength at outliers. The local weight  $\omega$  counteracts this imbalance by dividing by the total number of terms that a particular value  $u(x)$  appears in.

**4.3. Numerical results.** All experiments were performed on an Intel Xeon E5-2630 at 2.3 GHz with 64GB of RAM, MATLAB R2014a running on Scientific Linux 6.3, GCC 4.4.6, and Mosek 7. Run times were between several seconds for the geometric examples to several minutes for the full-size images, the majority of which was spent at the solution stage. The computation of the geodesic distances  $d_{\mathcal{M}}$  using the Fast Marching Method only took a few milliseconds in all cases, total preprocessing time including building the sparse matrix structures  $A$  and  $B$  took less than 5 seconds for the full-size images.

The solution of the Eikonal equation and computation of the weights as well as system matrix uses a custom C++ implementation. For solving the minimization problems we used the commercial Mosek solver with the CVX interface. Compared to the recent first-order approaches this allows to compute solutions with very high accuracy and therefore to evaluate the model without the risk of accidentally comparing only approximate solutions.

Figure 2 illustrates the effect of several classical local regularizers including TV, TV<sup>2</sup>, and TGV. As expected, TV generates the well-known staircasing effect, while TV<sup>2</sup> leads to oversmoothing of the jumps. TGV with hand-tuned parameters performs reasonably well, however it exhibits a characteristic pattern of clipping sharp local extrema. This behavior has also been analytically confirmed in [PB13, PS13] for the one-dimensional case.

Figure 3 shows the effect of our non-local Hessian-based method for the same problem. The piecewise affine signal is almost perfectly recovered. This is mostly due to the fact that the jumps are relatively large, which means that after computing the neighborhoods the circular region and the background are not coupled in terms of regularization. Therefore the regularization weight  $\alpha$  can be chosen very large, which results in virtually affine regions.

To see what happens with smaller jumps, we generated a pattern of opposing slopes (Figure 4). As expected both TGV as well as the non-local Hessian approach struggle when the jump is small. This shows the limitations of our approach for choosing the weights – while it adds some structural information to the

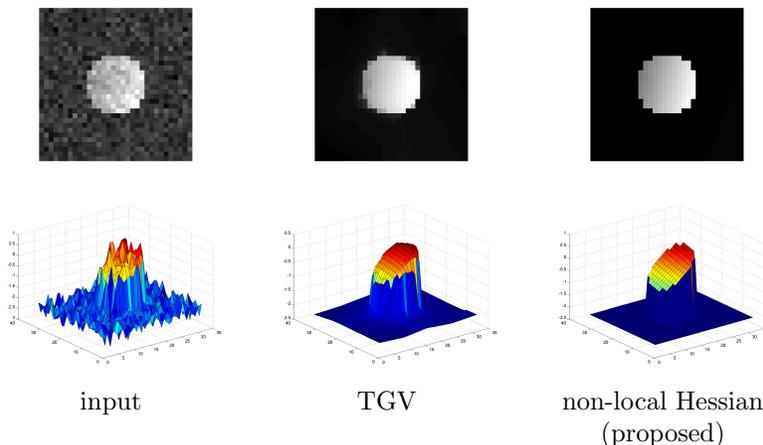


FIGURE 3. Non-local regularization of the problem in Figure 2. The adaptive choice of the neighbourhood and weights together with the non-local Hessian preserves the jumps, clip the top of the slope, and allows to perfectly reconstruct the piecewise affine signal.

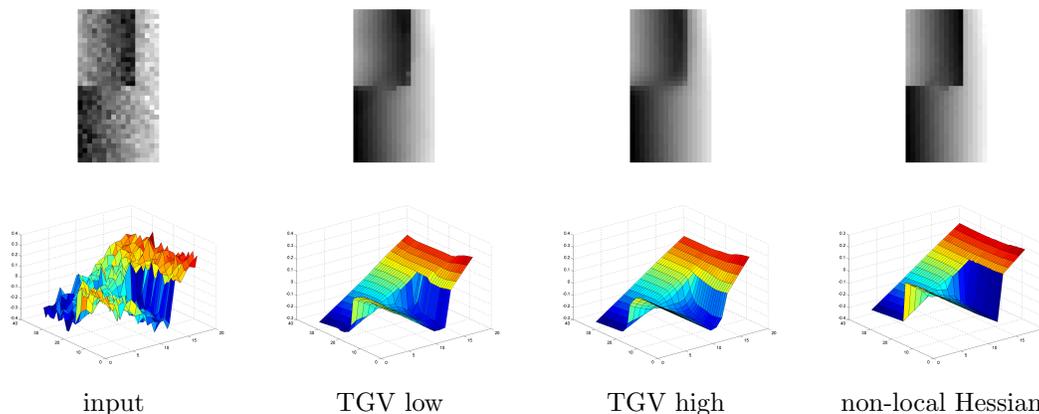


FIGURE 4. Denoising results for the “opposing slopes” image. The small jump at the crossing causes a slight amount of smoothing for both the TGV and non-local approaches. TGV with low regularization strength (TGV low,  $\alpha_0 = \alpha_1 = 0.8$ ) does reasonably well at preserving the strong jump on the right, but does not reconstruct the constant slope well. If one increases the regularization strength (TGV high,  $\alpha_0 = \alpha_1 = 1.5$ ), the jump is smoothed out. The non-local Hessian regularization ( $\alpha = 10^{-3}$ , note the different scales on the axes) fully preserves the large jump by design, and results in an overall cleaner reconstruction. The small amplitude differences at the crossing point cause a slight blur in both approaches.

regularizer, this information is still restricted to a certain neighbourhood of each point, and does not take into account the full global structure.

Figures 5 and 6 show a quantitative comparison of our non-local method with the results of the TGV approach and several classical regularizers. The parameters for each method were chosen by a grid search to yield the best PSNR. For all images we also provide a more realistic “structural similarity index” (SSIM) [WBSS04]. The non-local approach improves on TGV with respect to PSNR (34.09 vs. 33.28). However, it is interesting to look at the spatial distribution of the error: the parts where the non-local approach improves are exactly the local maxima, which are smoothed over by TGV. Surprisingly, this is hardly visible when looking at the images only (Figure 5), which is in line with the good results obtained with TGV for natural

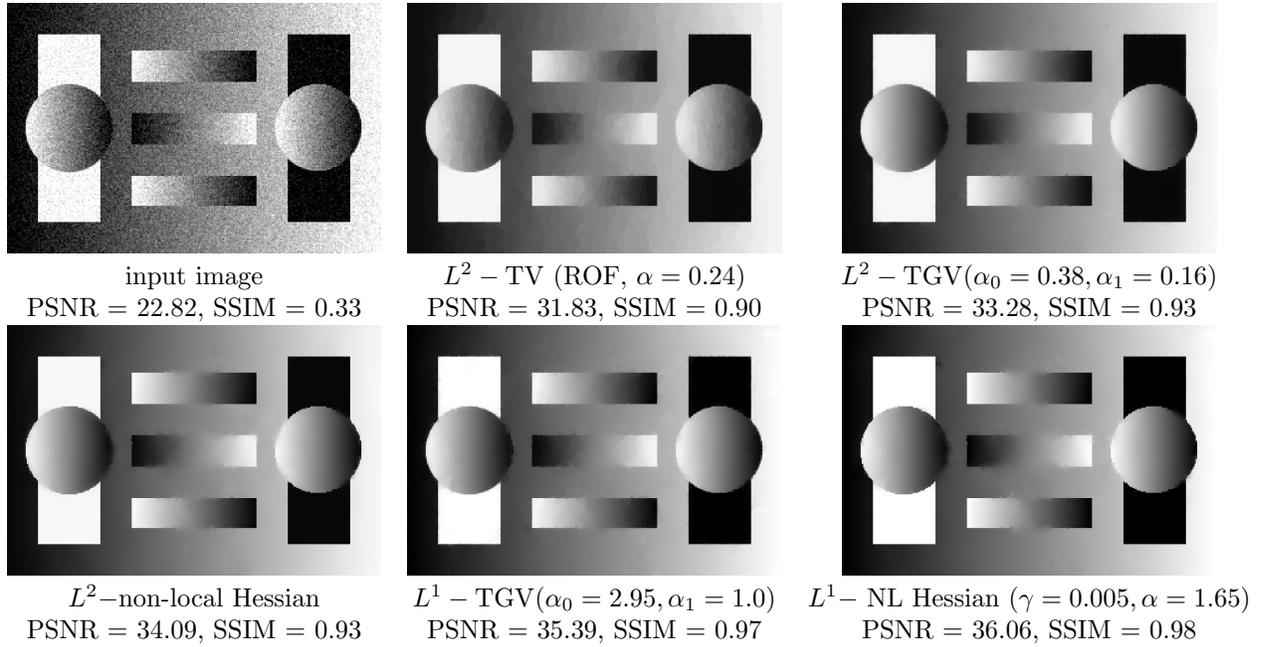


FIGURE 5. Denoising results for a geometric test image (see Figure 6 for the discussion).

images. However, this property might become a problem when the data is not a natural image, for example in the case of depth images or digital elevation maps. We refer to [LBL13] for a discussion of a problem that is more demanding in terms of the correct choice of the regularization.

A current limitation is that our choice of weights can result in a pixelated structure along slanted edges. This can happen when neighboring points are separated by a strong edge and therefore have no regularization in common. We conjecture that this effect could be overcome by ensuring that for neighbouring points  $x, y$  the weight  $\sigma_x(y)$  is always at least some positive constant, however we have not evaluated this direction further. Whether this is a desired behaviour or not depends on the underlying data – for applications such as segmentation a pixel-wise decision might actually be preferable.

Finally, Figure 8 shows an application to the “cameraman” image with  $L^2$  data term. Small details on the camera as well as long, thin structures such the camera handle and the highlights on the tripod are well preserved. In larger, more homogeneous areas the result is what would be expected from second-order smoothness.

In its basic form, our approach is data-driven, i.e., the weights are computed directly from the noisy data, while ideally they should be computed from the noise-free ground truth. We can approximate this *solution-driven* approach by iterating the whole process, each time recomputing the weights from the previous solution.

## 5. CONCLUSION

From the perspective of the analysis, the study of the non-local Hessian is a natural extension of the study of non-local gradients. While one has the straightforward observation we mention in the introduction – that one may use non-local gradients to characterize higher-order Sobolev spaces – the results of this paper alongside those of [MS14] provide a framework for general characterizations of Sobolev spaces of arbitrary (non-negative integer) order. This notion, that one can write down an integrated Taylor series and use this as a definition of a non-local differential object, is quite simple and yet leaves much to be explored. Let us mention several interesting questions to this effect that would add clarity to the picture that has developed thus far.

One of the foundations of the theory of Sobolev spaces is that of Sobolev inequalities. While for a related class of functionals, a Poincaré inequality has been established by Augusto Ponce [Pon04a], the lack of monotonicity of the non-local gradient has proven difficult in adapting his argument to our setting. This motivates the following open question.

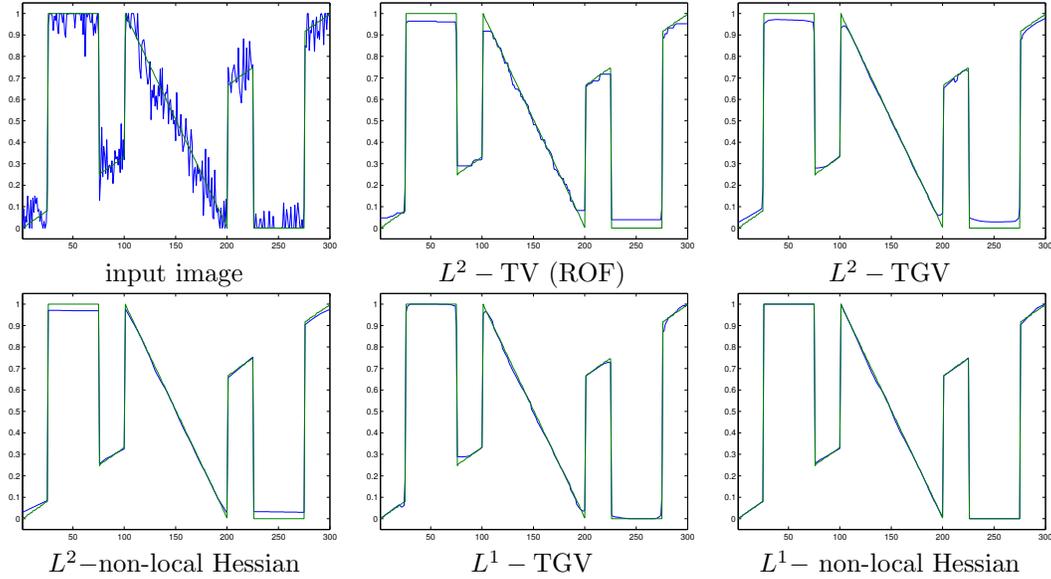


FIGURE 6. Slice at row 50 through the results images in Figure 5. The  $L^2$ -TV model shows staircasing as expected.  $L^2$ -TGV smoothes out sharp local extrema, which is preserved better by the non-local  $L^2$  model. Changing the data term to  $L^1$  additionally removes the slight contrast reduction, even though the original noisy data is clipped to the interval  $[0, 1]$ .

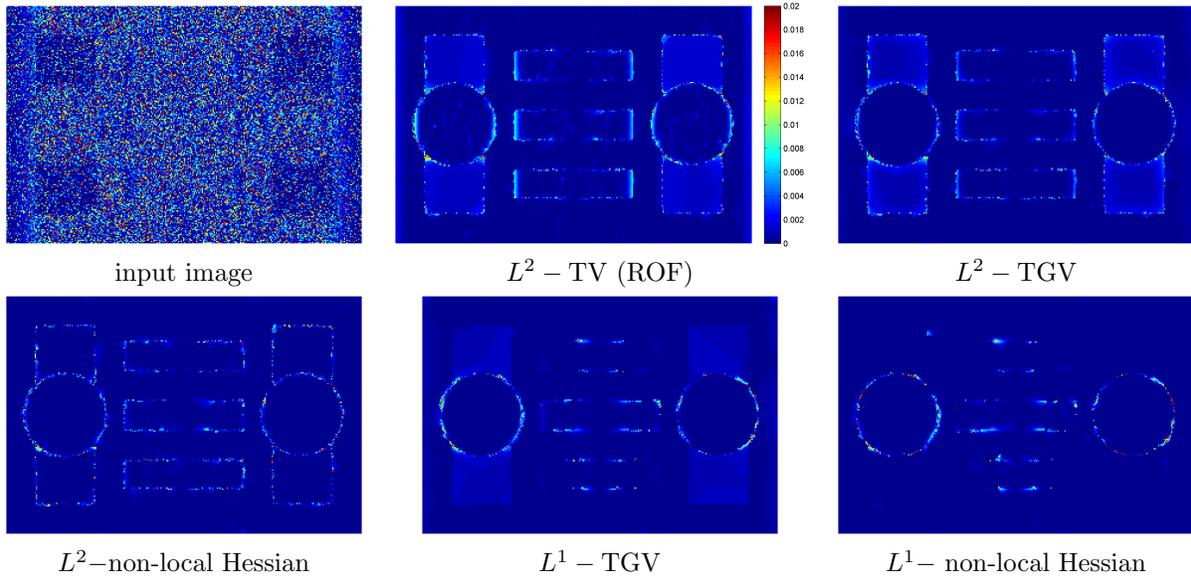


FIGURE 7. Distribution of the  $L^2$  error to the ground truth for the experiment in Figure 5. In the TGV case the errors spread out from the jumps. This is greatly reduced when using the non-local Hessian, as the cost of introducing a few more local errors.

**Open Question:** Can one find hypothesis on the approximation of the identity  $\rho_n$  so that there is a  $C > 0$  such that for all  $u$  in a suitable space one has the inequality

$$\int_{\Omega} |u - \mathcal{F}u|^p dx \leq C \int_{\Omega} |G_n u|^p dx,$$

for all  $n$  sufficiently large?

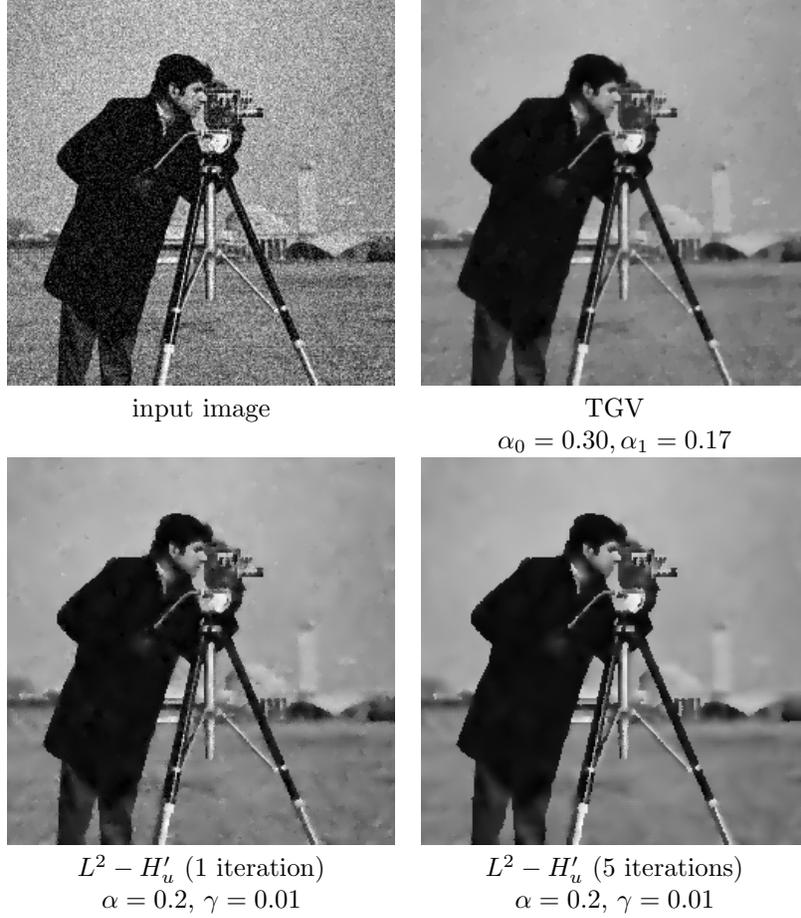


FIGURE 8. Non-local Hessian-based denoising with  $L^2$  data fidelity ( $p = 2$ ) on a real-world noisy input image (top left) with  $\sigma = 0.1$  Gaussian noise. Using a suitable choice of weights with a non-parametric neighborhood, small details and long, thin structures such as the camera handle can be preserved while removing noise in uniform structures (bottom left). By repeating the process several times, the result can be further improved (bottom right). Compared to TGV (upper right), edges are generally more pronounced, resulting in a slightly cartoon-like look.

In fact, beyond the multitude of questions one could explore by making a comparison of the results for non-local gradients and Hessians with known results in the Sobolev spaces, there are already several interesting questions for non-local gradients and Hessians in the regime  $p = 1$  that have not been satisfactorily understood from our perspective. For instance, in the first order setting, assuming  $u \in \text{BV}(\Omega)$ , one has the convergence of the total variations

$$\int_{\Omega} |G_n u| \, dx \rightarrow |Du|(\Omega).$$

While we have been able to show such a result for  $H_n^{ij}u$ ,  $i \neq j$ , we have not succeeded in demonstrating this in the case  $H_n^{ii}u$ , which if true would settle the following conjecture.

**Conjecture:** Suppose  $u \in \text{BV}^2(\mathbb{R}^N)$ . Then

$$\int_{\mathbb{R}^N} |H_n u| \, dx \rightarrow |D^2u|(\mathbb{R}^N).$$

This question is related to a larger issue needing clarification, which is that of the right assumptions for both the non-local gradient and non-local Hessian when  $p = 1$ . In both the paper [MS14] and this paper, the assumption utilized in the characterizations has been that the object is an  $L^1$  function. The more

natural assumption is that the non-local gradient or non-local Hessian exists as a measure, for which the same argument gives a characterization of  $BV(\Omega)$  or  $BV^2(\mathbb{R}^N)$ . However, the question of defining an integral functional of the non-local gradient or non-local Hessian is less clear. More understanding is required here as to the notion of integral functionals of a measure in the local case and the right framework to place the non-local objects within.

One can already see that from the analysis standpoint there are many interesting questions to explore in this area, while from the standpoint of applications we have seen that the relatively simple framework yields already quite interesting results. Here two interesting directions come into mind: developing strategies for choosing the weights, and moving on beyond second-order regularization.

While in the numerical section we outlined one potential useful choice for the weights and an application for the non-local Hessian – jump-preserving second-order regularization –, the approach is still somewhat local in the sense that we still restrict ourselves to a neighborhoods of each point. It would be interesting to find an application that allows to truly capitalize on the fact that we can include far-reaching interactions together with an underlying higher-order model.

It would also be interesting to see whether it is useful in applications to go to even higher orders than 2. While for natural images the use of such regularity is debatable, for other applications such as reconstructing digital elevation maps it can make the difference between success and failure [LBL13], and only requires replacing the quadratic model in the non-local Hessian (4.4) by a cubic- or even higher-order model.

#### APPENDIX

*Proof of Proposition 4.1* In the case  $N = 1$  we always have  $i_0 = j_0$  and verify that in fact both the integral and the right-hand side are equal to 2. For  $N = 2$  the integral becomes either

$$(5.1) \quad \int_0^{2\pi} \cos^2 \alpha \sin^2 \alpha d\alpha = \frac{\pi}{4} \quad \text{or} \quad \int_0^{2\pi} \cos^2 \alpha \cos^2 \alpha d\alpha = \frac{3\pi}{4},$$

which agrees with (4.5). For the general case  $N \geq 3$ , we apply the coarea formula in [AFP00, 2.93] using  $E = \mathcal{S}^{N-1}$ ,  $M = N$ ,  $N = N - 1$ ,  $k = 2$ ,  $f(x_1, x_2, \dots) = (x_1, x_2)$ , then

$$\begin{aligned} \int_{\mathcal{S}^{N-1}} g(x) C_k d^E f_x d\mathcal{H}^{N-1}(x) &= \int_{\mathbb{R}^2} \int_{\mathcal{S}^{N-1} \cap \{y_1=x_2, y_2=x_2\}} g(y) d\mathcal{H}^{N-3}(y) d(x_1, x_2) \\ &= \int_{B_1(0)} \int_{\sqrt{1-x_1^2-x_2^2} \mathcal{S}^{N-3}} g(x_1, x_2, x_r) d\mathcal{H}^{N-3}(x_r) d(x_1, x_2). \end{aligned}$$

In our case  $g(x) = x_1^2 x_2^2$ , thus  $g(x_1, x_2, x_r)$  is independent of  $x_r$  and we obtain

$$\int_{\mathcal{S}^{N-1}} x_1^2 x_2^2 C_k d^E f_x d\mathcal{H}^{N-1}(x) = \int_{B_1(0)} x_1^2 x_2^2 \int_{\sqrt{1-x_1^2-x_2^2} \mathcal{S}^{N-3}} d\mathcal{H}^{N-3}(x_r) d(x_1, x_2)$$

Consider  $C_k d^E f_x$ . For a given  $x \in \mathcal{S}^{N-1}$ , assume that  $B \in \mathbb{R}^N \times \mathbb{R}^{N-1}$  extends  $x$  to an orthonormal basis of  $\mathbb{R}^N$ . Then  $x + Bt$ ,  $t \in \mathbb{R}^{N-1}$ , parametrizes the tangent plane of  $E$  at  $x$ . The derivative of  $f(x + Bt) = (x_1, x_2)$  at  $x$  in direction  $t$  is

$$L := d^E f_x = \partial_t f(x + Bt)|_{t=0} = EB, \quad E := \begin{pmatrix} e_1^\top \\ e_2^\top \end{pmatrix}$$

We are interested in  $\sqrt{\det LL^\top} = \sqrt{\det(EBB^\top E)}$ . Now

$$I_2 = EI_n E^\top = E \begin{pmatrix} B & x \end{pmatrix} \begin{pmatrix} B^\top \\ x^\top \end{pmatrix} E^\top = EBB^\top E + Exx^\top E^\top,$$

thus

$$LL^\top = EBB^\top E = I_2 - Exx^\top E^\top = I_2 - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} x_1 & x_2 \end{pmatrix} = \begin{pmatrix} 1 - x_1^2 & -x_1 x_2 \\ -x_1 x_2 & 1 - x_2^2 \end{pmatrix},$$

and consequently

$$C_k d^E f_x = \sqrt{\det(LL^\top)} = \sqrt{(1 - x_1^2)(1 - x_2^2) - x_1^2 x_2^2} = \sqrt{1 - x_1^2 - x_2^2}.$$

Consequently

$$\int_{\mathcal{S}^{N-1}} x_1^2 x_2^2 (1 - x_1^2 - x_2^2) d\mathcal{H}^{N-1}(x) = \int_{B_1(0)} x_1^2 x_2^2 \int_{\sqrt{1-x_1^2-x_2^2}\mathcal{S}^{N-3}} d\mathcal{H}^{N-3}(x_r) d(x_1, x_2)$$

By putting the  $C_k$  term into the denominator of  $g$ , we obtain the more useful form

$$\int_{\mathcal{S}^{N-1}} x_1^2 x_2^2 d\mathcal{H}^{N-1}(x) = \int_{B_1(0)} \sqrt{1-x_1^2-x_2^2}^{-1} x_1^2 x_2^2 \int_{\sqrt{1-x_1^2-x_2^2}\mathcal{S}^{N-3}} d\mathcal{H}^{N-3}(x_r) d(x_1, x_2).$$

The right-hand side is

$$\begin{aligned} & \int_0^1 \int_0^{2\pi} r \sqrt{1-r^2}^{-1} r^4 \cos^2 \alpha \sin^2 \alpha \left| \sqrt{1-r^2} \mathcal{S}^{N-3} \right| dr d\alpha \\ &= \int_0^{2\pi} \cos^2 \alpha \sin^2 \alpha d\alpha \int_0^1 \sqrt{1-r^2} r^5 \left| \sqrt{1-r^2} \mathcal{S}^{N-3} \right| dr \\ &= \frac{\pi}{4} \int_0^1 \sqrt{1-r^2}^{-1} r^5 \left| \sqrt{1-r^2} \mathcal{S}^{N-3} \right| dr \\ &= \frac{\pi}{4} \int_0^1 \sqrt{1-r^2}^{-1} r^5 \sqrt{1-r^2}^{N-3} |\mathcal{S}^{N-3}| dr \\ &= \frac{\pi}{4} |\mathcal{S}^{N-3}| \int_0^1 r^5 \sqrt{1-r^2}^{N-4} dr \\ &= \frac{\pi}{4} |\mathcal{S}^{N-3}| \frac{8}{N(N+2)(N-2)} \\ &= \frac{\pi}{4} \frac{(N-2)\pi^{(N-2)/2}}{\Gamma(1+(N-2)/2)} \frac{8}{N(N+2)(N-2)} \\ &= \frac{2\pi^{N/2}}{N(N+2)\Gamma(N/2)}. \end{aligned}$$

Using the fact that  $\int_0^{2\pi} \cos^4(\alpha) d\alpha = 3 \int_0^{2\pi} \cos^2 \alpha \sin^2 \alpha d\alpha$  we finally obtain

$$C_{i_0 j_0} = \int_{\mathcal{S}^{N-1}} \nu_{i_0}^2 \nu_{j_0}^2 d\mathcal{H}^{N-1}(x) = \frac{|\mathcal{S}^{N-1}|}{N(N+2)} \cdot \begin{cases} 1, & i_0 \neq j_0, \\ 3, & i_0 = j_0. \end{cases}$$

This completes the proof.

## REFERENCES

- [AFP00] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, USA, 2000.
- [AK09] G. Aubert and P. Kornprobst, *Can the nonlocal characterization of Sobolev spaces by Bourgain et al. be useful for solving variational problems?*, *SIAM Journal on Numerical Analysis* **47** (2009), no. 2, 844–860, <http://dx.doi.org/10.1137/070696751>.
- [BBM01] J. Bourgain, H. Brezis, and P. Mironescu, *Another look at Sobolev spaces*, *Optimal Control and Partial Differential Equations. A Volume in Honour of A. Bensoussan's 60th Birthday* (2001), 439–455.
- [BCM05] A. Buades, B. Coll, and J.M. Morel, *A review of image denoising algorithms, with a new one*, *Multiscale Modeling & Simulation* **4** (2005), no. 2, 490–530, <http://dx.doi.org/10.1137/040616024>.
- [BEG08] A. Bertozzi, S. Esedoglu, and A. Gillette, *Analysis of a two-scale Cahn-Hilliard model for binary image inpainting*, *Multiscale Modeling and Simulation* **6** (2008), no. 3, 913–936, <http://dx.doi.org/10.1137/060660631>.
- [BG04] Andrea L. Bertozzi and John B. Greer, *Low-curvature image simplifiers: Global regularity of smooth solutions and Laplacian limiting schemes*, *Communications on Pure and Applied Mathematics* **57** (2004), no. 6, 764–790.
- [BHS09] M. Burger, L. He, and C.B. Schönlieb, *Cahn-Hilliard inpainting and a generalization for grayvalue images*, *SIAM Journal on Imaging Sciences* **2** (2009), no. 4, 1129–1167, <http://dx.doi.org/10.1137/080728548>.
- [BKP10] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, *SIAM Journal on Imaging Sciences* **3** (2010), no. 3, 492–526, <http://dx.doi.org/10.1137/090769521>.
- [BP10] M. Bergounioux and L. Piffet, *A second-order model for image denoising*, *Set-Valued and Variational Analysis* **18** (2010), no. 3-4, 277–306, <http://dx.doi.org/10.1007/s11228-010-0156-6>.
- [Bra02] A. Braides,  *$\Gamma$ -convergence for beginners*, Oxford lecture series in mathematics and its applications, 2002.
- [Bre83] Haïm Brezis, *Analyse fonctionnelle*, vol. 5, Masson, 1983.
- [BV11] K. Bredies and T. Valkonen, *Inverse problems with second-order total generalized variation constraints*, *Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications*, Singapore, 2011.

- [CEP07] T.F. Chan, S. Esedoglu, and F.E. Park, *Image decomposition combining staircase reduction and texture extraction*, Journal of Visual Communication and Image Representation **18** (2007), no. 6, 464–486, <http://dx.doi.org/10.1016/j.jvcir.2006.12.004>.
- [CKS02] T.F. Chan, S.H. Kang, and J. Shen, *Euler's elastica and curvature-based inpainting*, SIAM Journal on Applied Mathematics **63** (2002), no. 2, 564–592, <http://dx.doi.org/10.1137/S0036139901390088>.
- [CL97] A. Chambolle and P.L. Lions, *Image recovery via total variation minimization and related problems*, Numerische Mathematik **76** (1997), 167–188, <http://dx.doi.org/10.1007/s002110050258>.
- [CMM01] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM Journal on Scientific Computing **22** (2001), no. 2, 503–516, <http://dx.doi.org/10.1137/S1064827598344169>.
- [CS01] T.F. Chan and J. Shen, *Nontexture inpainting by curvature-driven diffusions*, Journal of Visual Communication and Image Representation **12** (2001), no. 4, 436–449, <http://dx.doi.org/10.1006/jvci.2001.0487>.
- [Dem85] F. Demengel, *Fonctions à Hessien borné*, Annales de l'Institut Fourier **34** (1985), 155–190.
- [DM93] G. Dal Maso, *Introduction to  $\Gamma$ -convergence*, Birkhauser, 1993.
- [DMFLM09] G. Dal Maso, I. Fonseca, G. Leoni, and M. Morini, *A higher order model for image restoration: the one dimensional case*, SIAM Journal on Mathematical Analysis **40** (2009), no. 6, 2351–2391, <http://dx.doi.org/10.1137/070697823>.
- [DWB09] S. Didas, J. Weickert, and B. Burgeth, *Properties of higher order nonlinear diffusion filtering*, J. Math. Imaging Vis. **35** (2009), 208–226.
- [GO08] G. Gilboa and S. Osher, *Nonlocal operators with applications to image processing*, Multiscale Modeling & Simulation **7** (2008), no. 3, 1005–1028, <http://dx.doi.org/10.1137/070698592>.
- [HS06] W. Hinterberger and O. Scherzer, *Variational methods on the space of functions of bounded Hessian for convexification and denoising*, Computing **76** (2006), no. 1-2, 109–133, <http://dx.doi.org/10.1007/s00607-005-0119-1>.
- [KOJ05] S. Kindermann, S. Osher, and P. Jones, *Deblurring and denoising of images by nonlocal functionals*, Multiscale Modeling & Simulation **4** (2005), no. 4, 1091–1115, <http://dx.doi.org/10.1137/050622249>.
- [LBL13] F. Lenzen, F. Becker, and J. Lellmann, *Adaptive second-order total variation: An approach aware of slope discontinuities*, Scale Space and Variational Methods in Computer Vision, Lecture Notes in Computer Science, vol. 7893, Springer Berlin Heidelberg, 2013, [http://dx.doi.org/10.1007/978-3-642-38267-3\\_6](http://dx.doi.org/10.1007/978-3-642-38267-3_6), pp. 61–73.
- [LBU12] S. Lefkimmiatis, A. Bourquard, and M. Unser, *Hessian-based norm regularization for image restoration with biomedical applications*, IEEE Transactions on Image Processing **21** (2012), 983–995, <http://dx.doi.org/10.1109/TIP.2011.2168232>.
- [LDM07] R. Lerallut, E. Decencière, and F. Meyer, *Image filtering using morphological amoebas*, Image and Vision Computing **25** (2007), no. 4, 395–404, <http://dx.doi.org/10.1016/j.imavis.2006.04.018>.
- [LLT03] M. Lysaker, A. Lundervold, and X.C. Tai, *Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time*, IEEE Transactions on Image Processing **12** (2003), no. 12, 1579–1590, <http://dx.doi.org/10.1109/TIP.2003.819229>.
- [LS11] G. Leoni and D. Spector, *Characterization of Sobolev and BV spaces*, Journal of Functional Analysis **261** (2011), no. 10, 2926–2958, <http://dx.doi.org/10.1016/j.jfa.2011.07.018>.
- [LS14] ———, *Corrigendum to Characterization of Sobolev and BV spaces [J. Funct. Anal. 261 (10) (2011) 2926–2958]*, Journal of Functional Analysis **266** (2014), no. 2, 1106–1114, <http://dx.doi.org/10.1016/j.jfa.2013.10.026>.
- [LT06] M. Lysaker and X.C. Tai, *Iterative image restoration combining total variation minimization and a second-order functional*, International Journal of Computer Vision **66** (2006), no. 1, 5–18, <http://dx.doi.org/10.1007/s11263-005-3219-7>.
- [LTC13] R. Lai, X. Tai, and T. Chan, *A ridge and corner preserving model for surface restoration*, SIAM Journal on Scientific Computing **35** (2013), no. 2, A675–A695, <http://dx.doi.org/10.1137/110846634>.
- [Men12] T. Mengesha, *Nonlocal Korn-type characterization of Sobolev vector fields*, Communications in Contemporary Mathematics **14** (2012), no. 4, <http://dx.doi.org/10.1142/S0219199712500289>.
- [MM98] S. Masnou and J.M. Morel, *Level lines based disocclusion*, IEEE Proceedings of International Conference on Image Processing, 1998, <http://dx.doi.org/10.1109/ICIP.1998.999016>, pp. 259–263.
- [MS14] T. Mengesha and D. Spector, *Localization of nonlocal gradients in various topologies*, Calculus of Variations and Partial Differential Equations (2014).
- [Nir59] L. Nirenberg, *On elliptic partial differential equations*, Annali della Scuola Normale Superiore di Pisa, 3<sup>e</sup> serie **13** (1959), no. 2, 115–162.
- [OF03] S. Osher and R. Fedkiw, *Level set methods and dynamic implicit surfaces*, Springer, 2003.
- [PB13] K. Papafitsoros and K. Bredies, *A study of the one dimensional total generalised variation regularisation problem*, submitted (2013), <http://arxiv.org/abs/1309.5900>.
- [Pon04a] A.C. Ponce, *An estimate in the spirit of Poincaré's inequality*, Journal of European Mathematical Society **6** (2004), no. 1, 1–15, <http://dx.doi.org/10.4171/JEMS/1>.
- [Pon04b] ———, *A new approach to Sobolev spaces and connections to  $\Gamma$ -convergence*, Calculus of Variations and Partial Differential Equations **19** (2004), no. 3, 229–255, <http://dx.doi.org/10.1007/s00526-003-0195-z>.
- [PS08] C. Pöschl and O. Scherzer, *Characterization of minimisers of convex regularization functionals*, Contemporary mathematics **451** (2008), 219–248.
- [PS13] ———, *Exact solutions of one-dimensional TGV*, arXiv preprint (2013), <http://arxiv.org/abs/1309.7152>.
- [PS14] K. Papafitsoros and C.B. Schönlieb, *A combined first and second order variational approach for image reconstruction*, Journal of Mathematical Imaging and Vision **48** (2014), no. 2, 308–338, <http://dx.doi.org/10.1007/s10851-013-0445-4>.

- [PSS13] K. Papafitsoros, C.B. Schönlieb, and B. Sengul, *Combined First and Second Order Total Variation inpainting using Split Bregman*, Image Processing On Line **2013** (2013), 112–136, <http://dx.doi.org/10.5201/ipol.2013.40>.
- [RBP14] R. Ranftl, K. Bredies, and T. Pock, *Non-local total generalized variation for optical flow estimation*, Proceedings of the International Conference on Computer Vision and Applications, Springer, 2014, [http://dx.doi.org/10.1007/978-3-319-10590-1\\_29](http://dx.doi.org/10.1007/978-3-319-10590-1_29), pp. 439–454.
- [Rin00] W. Ring, *Structural properties of solutions to total variation regularization problems*, ESAIM: Mathematical Modelling and Numerical Analysis **34** (2000), no. 4, 799–810, <http://dx.doi.org/10.1051/m2an:2000104>.
- [ROF92] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Non-linear Phenomena **60** (1992), no. 1-4, 259–268, [http://dx.doi.org/10.1016/0167-2789\(92\)90242-F](http://dx.doi.org/10.1016/0167-2789(92)90242-F).
- [Sch98] O. Scherzer, *Denosing with higher order derivatives of bounded variation and an application to parameter estimation*, Computing **60** (1998), no. 1, 1–27, <http://dx.doi.org/10.1007/BF02684327>.
- [Set99] J.A. Sethian, *Level set methods and fast marching methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, vol. 3, Cambridge University Press, 1999.
- [SS08] S. Setzer and G. Steidl, *Variational methods with higher order derivatives in image processing*, Approximation XII (2008), 360–386.
- [SS14] T. Shieh and D. Spector, *On a new class of fractional partial differential equations*, Advances in Calculus of Variations (2014), <http://dx.doi.org/10.1515/acv-2014-0009>.
- [SST11] S. Setzer, G. Steidl, and T. Teuber, *Infimal convolution regularizations with discrete  $\ell_1$ -type functionals*, Communications in Mathematical Sciences **9** (2011), 797–872, <http://dx.doi.org/10.4310/CMS.2011.v9.n3.a7>.
- [WBSS04] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Transactions on Image Processing **13** (2004), no. 4, 600–612, <http://dx.doi.org/10.1109/TIP.2003.819861>.
- [WBV11] M. Welk, M. Breuß, and O. Vogel, *Morphological amoebas are self-snakes*, Journal of Mathematical Imaging and Vision **39** (2011), no. 2, 87–99, <http://dx.doi.org/10.1007/s10851-010-0228-0>.
- [Wel12] M. Welk, *Amoeba active contours*, Scale Space and Variational Methods in Computer Vision, vol. 6667, 2012, [http://dx.doi.org/10.1007/978-3-642-24785-9\\_32](http://dx.doi.org/10.1007/978-3-642-24785-9_32), pp. 374–385.

(Jan Lellmann, Konstantinos Papafitsoros, Carola Schönlieb) DEPARTMENT OF APPLIED MATHEMATICS AND THEORETICAL PHYSICS, UNIVERSITY OF CAMBRIDGE, UNITED KINGDOM

(Daniel Spector) TECHNION, ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA, ISRAEL AND DEPARTMENT OF APPLIED MATHEMATICS, NATIONAL CHIAO TUNG UNIVERSITY, HSINCHU, TAIWAN

*E-mail address:* J.Lellmann@damtp.cam.ac.uk

*E-mail address:* k.papafitsoros@maths.cam.ac.uk

*E-mail address:* C.B.Schoenlieb@damtp.cam.ac.uk

*E-mail address:* dspector@tx.technion.ac.il