

## Analysis and Application of a Nonlocal Hessian\*

Jan Lellmann<sup>†</sup>, Konstantinos Papafitsoros<sup>†</sup>, Carola Schönlieb<sup>†</sup>, and Daniel Spector<sup>‡</sup>

**Abstract.** In this work we introduce a formulation for a nonlocal Hessian that combines the ideas of *higher-order* and *nonlocal* regularization for image restoration, extending the idea of nonlocal gradients to higher-order derivatives. By intelligently choosing the weights, the model allows us to improve on the current state of the art higher-order method, total generalized variation, with respect to overall quality and preservation of 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 it allows a novel characterization of higher-order Sobolev and BV spaces.

**Key words.** nonlocal Hessian, nonlocal total variation regularization, variational methods, fast marching method, amoeba filters

**AMS subject classifications.** 65D18, 68U10, 94A08, 35A15, 49J40, 49Q20, 26B30, 26B35, 46E35

**DOI.** 10.1137/140993818

**1. Introduction and context.** 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  $Du$  (the distributional derivative of  $u$ , which has finite total mass when one assumes  $u$  is of bounded variation [AFP00]). Among the known defects of the model is the staircasing effect, where affine portions of the image are replaced by flat regions and newly created artificial boundaries, stemming from the use of the TV term in regularization. It is then natural to investigate the replacement of the total variation with

\*Received by the editors October 31, 2014; accepted for publication (in revised form) July 10, 2015; published electronically October 6, 2015. This project was supported by King Abdullah University of Science and Technology (KAUST) award KUK-I1-007-43.

<http://www.siam.org/journals/siims/8-4/99381.html>

<sup>†</sup>Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, United Kingdom (j.lellmann@damtp.cam.ac.uk, kp366@cam.ac.uk, cbs31@cam.ac.uk). The first author's research was supported by Leverhulme Early Career Fellowship ECF-2013-436. The second author's research was supported by the Cambridge Centre for Analysis and the EPSRC. The third author's research was partially supported by EPSRC grants EP/J009539/1 and EP/M00483X/1.

<sup>‡</sup>Technion, Israel Institute of Technology, Haifa, Israel, and Department of Applied Mathematics, National Chiao Tung University, Hsinchu 30010, Taiwan (dspector@math.nctu.edu.tw). This author's research was supported in part by a Technion Fellowship and by Taiwan Ministry of Science and Technology research grant 103-2115-M-009-016-MY2.

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 nonlocal term (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 nonlocal—a *nonlocal Hessian*—and utilize it in a model for image restoration. Our numerical experiments demonstrate that using this regularization with a suitable choice of weights enables us to derive specialized models that compete with current state of the art higher-order methods such as the total generalized variation [BKP10]. Meanwhile, our analysis justifies the nomenclature nonlocal Hessian through its connection with recent work on nonlocal gradients [MS15]. In particular, we perform rigorous localization analysis which parallels the first-order case.

**Background on higher-order regularization.** The use of nonsmooth regularization terms such as the total variation in image reconstruction results in a nonlinear smoothing of reconstructed images. As a consequence, one observes a greater degree of smoothing in homogeneous areas of the image domain while preserving characteristic structures such as edges. In particular, total variation regularization performs well if the reconstructed image is piecewise constant. The drawback of such a regularization procedure becomes apparent as soon as images or signals (in one dimension) are considered which not only consist of flat regions and jumps but also possess slanted regions, i.e., piecewise linear parts. The artifact introduced by total variation regularization in this case is called staircasing. One possible approach to improve total variation minimization is the introduction of higher-order derivatives in the regularizer, whose literature we now briefly review.

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

$$(1.1) \quad \min_{(u_1, u_2)} \frac{1}{2} \int_{\Omega} (u_1 + u_2 - g)^2 dx + \alpha \text{TV}(u_1) + \beta \text{TV}^2(u_2),$$

where  $\text{TV}^2(u_2) := |D^2 u_2|(\Omega)$  is the total variation of the distributional Hessian of  $u_2$ . Then,  $u_1$  and  $u_2$  are the piecewise constant and the 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 combining first and second regularization originate, for instance, from Chan, Marquina, and Mulet [CMM01], who consider total variation minimization together with weighted versions of the Laplacian, the Euler- elastica functional [MM98, CKS02], which combines total variation regularization with curvature penalization, and many more [LT06, LTC13, PS14, PSS13, Ber14]. Recently, Bredies, Kunisch, and Pock 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

$$(1.2) \quad \text{TGV}_{\alpha}^k(u) = \sup \left\{ \int_{\Omega} u \operatorname{div}^k \xi \, dx : \xi \in C_c^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^N)), \|\operatorname{div}^l \xi\|_{\infty} \leq \alpha_l, l = 0, \dots, k-1 \right\},$$

where  $\operatorname{Sym}^k(\mathbb{R}^N)$  denotes the space of symmetric tensors of order  $k$  with arguments in  $\mathbb{R}^N$ ,

and  $\alpha_l$  are fixed positive parameters. For the case  $k = 2$ , its formulation for the solution of general inverse problems was given in [BV11].

The idea of pure bounded Hessian regularization is considered by Lysaker, Lundervold, and Tai [LLT03], Scherzer [Sch98], Hinterberger and Scherzer [HS06], Lefkimmatis, Bourquard, and Unser [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, Esedoglu, and Park 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  $(l-1)$ st 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 nonquadratic functions  $f$ . There are also works on higher-order PDE methods for image regularization; see, 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 staircasing inherent to total variation [Rin00].

**Higher-order nonlocal regularization.** One possible approach to a higher-order extension of nonlocal regularization has been proposed recently in the work [RBP14], with optical flow being the main application. The authors start with the cascading formulation of (second-order) TGV,

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

which reduces the higher-order differential operators that appear in the definition of TGV to a special type of infimal convolution of two terms involving only first-order derivatives [BV11]. These can then be replaced by classical first-order nonlocal derivatives, and one obtains an energy of the form

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

This formulation takes into account the higher-order differential information via the second term in the minimization, and the weighting parameters  $\alpha_0$  and  $\alpha_1$  are now spatially dependent. Even though this approach can be adapted for other imaging tasks, e.g., denoising, it is not clear how to choose these weighting functions.

In this paper we define a different type of higher-order nonlocal regularizer, providing as well a rule for choosing the corresponding weighting functions for optimal results. Before we proceed we recall some basic facts about nonlocal gradients.

**Background on nonlocal gradients.** In the first-order setting, the analysis of nonlocal gradients and their associated energies finds its origins in the 2001 paper of Bourgain, Brezis, and Mironescu [BBM01]. In their paper, Bourgain, Brezis, and Mironescu introduce energies of the form

$$(1.3) \quad F_n u := \int_{\Omega} \left( \int_{\Omega} \frac{|u(x) - u(y)|^{pq}}{|x - y|^{pq}} \rho_n(x - y) dx \right)^{\frac{1}{q}} dy,$$

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  $n \in \mathbb{N}$ :

$$(1.4) \quad \rho_n(x) \geq 0,$$

$$(1.5) \quad \int_{\mathbb{R}^N} \rho_n(x) dx = 1,$$

$$(1.6) \quad \lim_{n \rightarrow \infty} \int_{|x| > \gamma} \rho_n(x) dx = 0 \quad \forall \gamma > 0.$$

An example of such a family of mollifiers are the standard Gaussian kernels that converge to a Dirac  $\delta$  as  $n$  tends to infinity. Let us here remark that a perhaps more appropriate terminology for these functionals in image processing is *semilocal*, since asymptotically there is no possibility of nonlocality, in contrast to the genuine nonlocality allowed in image processing.

The work [BBM01] connects the finiteness of the limit as  $n \rightarrow \infty$  of the functional (1.3) 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 introduce an energy similar to (1.3), terming it a nonlocal total variation, while the connection of the two and the introduction of the parameter  $q$  is due to Leoni and Spector [LS11]. In particular, they show in [LS14] that for  $p = 1$  the functionals (1.3)  $\Gamma$ -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 nonlocal total variations, and for our purposes here it will be useful to consider the second. This alternative involves introducing a nonlocal gradient operator, defined by

$$(1.7) \quad G_n u(x) := N \int_{\Omega} \frac{u(x) - u(y)}{|x - y|} \frac{x - y}{|x - y|} \rho_n(x - y) dy, \quad x \in \Omega,$$

for  $u \in C_c^1(\Omega)$ . Then, one defines the nonlocal total variation as the  $L^1$  norm of (1.7). The localization analysis of the nonlocal gradient (1.7) has been performed by Mengesha and

Spector in [MS15], where a more general (and technical) distributional definition is utilized. Their first observation is that the definition of the nonlocal gradient via the Lebesgue integral (1.7) extends to spaces of weakly differentiable functions. In this regime they discuss the localization of (1.7). They prove that the nonlocal gradient converges to its local analogue  $\nabla u$  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 nonlocal total variation energies defined via nonlocal gradients of the form (1.7) to the local total variation.

One way to extend the results of Mengesha and Spector to the higher-order case is to simply study the functional that results after substituting  $u$  with  $\nabla u$  in (1.7). Then a nonlocal Hessian could be defined via

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

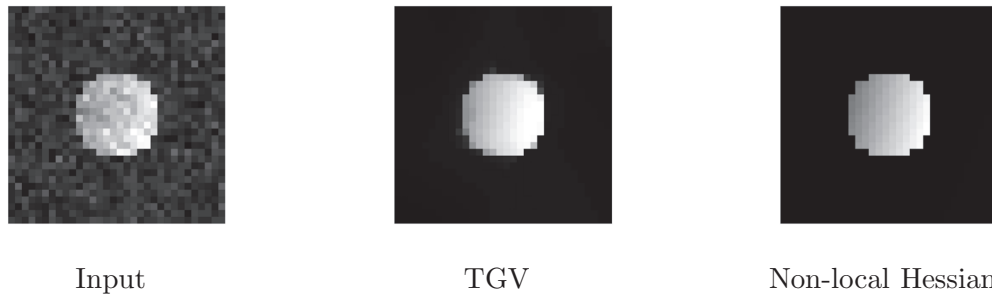
where  $\otimes$  denotes the standard tensor multiplication of vectors. While one can obtain some straightforward characterization of  $W^{2,p}(\mathbb{R}^N)$  and  $BV^2(\mathbb{R}^N)$  in this way, we find it advantageous to utilize a nonlocal Hessian that is derivative-free and therefore pursue an alternative approach.

**A nonlocal Hessian tuned for imaging tasks.** We define an *implicit* nonlocal 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*:

$$(1.9) \quad (G'_u(x), H'_u(x)) = \underset{G_u \in \mathbb{R}^N, H_u \in \text{Sym}(\mathbb{R}^{N \times N})}{\text{argmin}} \frac{1}{2} \int_{\Omega - \{x\}} \left( u(x + z) - u(x) - G_u^\top z - \frac{1}{2} z^\top H_u z \right)^2 \sigma_x(z) dz,$$

where  $\Omega - \{x\} = \{y - x : y \in \Omega\}$  and  $\sigma_x$  is an appropriate weight function for each  $x \in \Omega$ . Such a definition has the advantage of the freedom to choose the weights  $\sigma_x$  as one sees fit. Of primary interest to our work are two questions: How does the nonlocal Hessian perform in comparison to the known state of the art methods? And in what way is it connected to the classical Hessian?

To answer the first question, the model depends on the choice of weights, and of practical relevance is the question of how to choose them for a particular purpose. The first point to mention in this regard is that as the objectives of the minimization problems (1.9) are quadratic, their solutions can be characterized by linear optimality conditions. Thus functionals based on the implicit nonlocal 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. In particular, in this work we develop a method of choosing weights to construct a regularizer that both favors piecewise affine functions while allowing for jumps in the data. Our motivation stems from the recent discussion of “amoeba” filters in [LDM07, WBV11, Wel12], which combine standard filters such as median filters with nonparametric structuring elements that are based on the data; that is, in long thin objects they would extend along the structure and thus



**Figure 1.** Illustration of the capability of the proposed nonlocal Hessian regularization to obtain true piecewise affine reconstructions in a denoising example.

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. In a similar manner, we utilize the geodesic distance to set the weights  $\sigma_x$ . This allows us to get a very close approximation to true piecewise affine regularization, in many cases improving on the results obtained using TGV; see Figure 1 for a proof of concept. We present several experiments in section 4.3 that show the performance of this choice against the state of the art.

As to the second question, in the general form of (1.9), the problem is considerably harder to treat analytically, and so we will restrict ourselves to the special case of radial weights. In particular, assuming some mild regularity assumptions on  $u$  and considering the problem (1.9) with weights  $\rho_n(z)/|z|^4$ ,

$$(1.10) \quad (G'_u(x), H'_u(x)) = \underset{G_u \in \mathbb{R}^N, H_u \in \text{Sym}(\mathbb{R}^N \times \mathbb{R}^N)}{\text{argmin}} \frac{1}{2} \int_{\mathbb{R}^N} \left( u(x+z) - u(x) - G_u^\top z - \frac{1}{2} z^\top H_u z \right)^2 \frac{\rho_n(z)}{|z|^4} dz,$$

we will show in Theorem 4.1 that  $H'_u$  agrees with the following natural explicit definition of nonlocal Hessian.

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

$$(1.11) \quad H_n u(x) := \frac{N(N+2)}{2} \int_{\mathbb{R}^N} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \frac{\left( z \otimes z - \frac{|z|^2}{N+2} I_N \right)}{|z|^2} \rho_n(z) dz, \quad x \in \mathbb{R}^N,$$

where here  $I_N$  is the  $N \times N$  identity matrix and  $\rho_n$  is a sequence satisfying (1.4)–(1.6).

We note here that the presence of the constant  $N(N+2)/2$  as well as the term  $z \otimes z - \frac{|z|^2}{N+2} I_N$  ensure that (1.11) has the right localization properties; see section 3 for more details. The assertion of Theorem 4.1 is that with the preceding choice of weights one has the equivalence

$$H'_u(x) \equiv \frac{N(N+2)}{2} \int_{\mathbb{R}^N} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \frac{\left( z \otimes z - \frac{|z|^2}{N+2} I_N \right)}{|z|^2} \rho_n(z) dz.$$

**Results concerning the explicit nonlocal Hessian.** From the standpoint of analysis, the explicit version of nonlocal Hessian (1.11) is more natural, for which we are able to prove a number of results analogous to the first-order case studied by Mengesha and Spector [MS15]. Let us first extend the definition to functions which are not necessarily smooth and compactly supported, as typical for operators acting on spaces of weakly differentiable functions.

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

$$(1.12) \quad \langle \mathfrak{H}_n^{ij} u, \varphi \rangle := \int_{\mathbb{R}^N} u H_n^{ij} \varphi \, dx$$

for  $\varphi \in C_c^\infty(\mathbb{R}^N)$ , where  $H_n^{ij} \phi$  denotes the  $i, j$ th element of the nonlocal Hessian matrix (1.11).

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 (nonlocal integration by parts).** Suppose that  $u \in L^p(\mathbb{R}^N)$  for some  $1 \leq p < +\infty$  and  $\frac{|u(x+z) - 2u(x) + u(x-z)|^q}{|z|^{2q}} \rho_n(z) \in L^1(\mathbb{R}^N \times \mathbb{R}^N)$  for some  $1 \leq q \leq +\infty$ . Then the distribution  $\mathfrak{H}_n u$  can be represented by the function  $H_n u$ , i.e., for any  $\varphi \in C_c^2(\mathbb{R}^N)$  and  $i, j = 1, \dots, N$ ,

$$(1.13) \quad \langle \mathfrak{H}_n^{ij} u, \varphi \rangle = \int_{\mathbb{R}^N} H_n^{ij} u(x) \varphi(x) \, dx,$$

and also  $H_n u \in L^1(\mathbb{R}^N, \mathbb{R}^{N \times N})$ .

We will see in section 3, in Lemmas 3.1 and 3.4, 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.

Then the main analysis we undertake in this paper are the following results, 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 nonlocal Hessian.

**Theorem 1.6.** Let  $u \in BV^2(\mathbb{R}^N)$  and  $\mu_n := H_n \mathcal{L}^N$  be a sequence of  $\mathbb{R}^{N \times N}$ -valued measures. Then

$$\mu_n \rightarrow D^2 u, \quad \text{weakly}^* \text{ in the space of Radon measures};$$

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

$$(1.14) \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^N} H_n u(x) \cdot \phi(x) dx = \int_{\mathbb{R}^N} \phi(x) \cdot dD^2 u.$$

We have seen that the nonlocal 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  $u \in \text{BV}^2(\mathbb{R}^N)$  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*

$$(1.15) \quad u \in W^{2,p}(\mathbb{R}^N) \iff \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^N} |\mathfrak{H}_n u(x)|^p dx < \infty.$$

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

$$(1.16) \quad u \in \text{BV}^2(\mathbb{R}^N) \iff \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^N} |\mathfrak{H}_n u(x)| dx < \infty.$$

Note that when we write  $\int_{\mathbb{R}^N} |\mathfrak{H}_n u(x)|^p dx$  we mean that the distribution  $\mathfrak{H}_n u$  is representable by an  $L^p$  function.

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 nonlocal Hessian energies to the energy of the Hessian.

**Theorem 1.8.** *Let  $u \in L^1(\mathbb{R}^N)$ . Then*

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

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, combined with an equicoercivity assumption, implies convergence of the minimizers in a suitable topology [Bra02, Chap. 1.5]. Assuming equicoercivity, Theorem 1.8 then guarantees that under a suitable choice of weights, the solutions of a class of *nonlocal* Hessian-based problems converges to the solution of the *local* Hessian-regularized problem, and thus our notion of “nonlocal Hessian” is justified. Note that because Theorem 4.1 connects the implicit and explicit definitions of nonlocal Hessian, these results equivalently read that for radial weights that concentrate to a Dirac mass our nonlocal Hessian experiments concentrate to the bounded Hessian framework.

**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 nonlocal Hessian functional (1.11). After a justification of the introduction of its distributional form, we proceed in section 3.1 with the localization of (1.11) to the classical Hessian for smooth



functions  $u$ . The localization of (1.11) 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 nonlocal 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 nonlocal Hessian (1.9), 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 (1.9) 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 domain of the function, while the second denotes its range. Whenever the range is omitted, it is assumed that the functions are real valued. When a function space is in the subscript of a norm, only the domain is specified for the sake of better readability.

We use  $dx, dy, dz$  for various integrations with respect to 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 nonlocal Hessian when this is a function. As we have already seen,  $\mathfrak{H}$  denotes the distributional form of the nonlocal 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 analog result to the strict approximation by smooth functions for the classical BV case; see [AFP00].

**Theorem 2.1 (BV<sup>2</sup> strict approximation by smooth functions [Dem85]).** *Let  $\Omega \subseteq \mathbb{R}^N$  be open, and let  $u \in BV^2(\Omega)$ . 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,*

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

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 componentwise.

We now remind the reader about some basic facts concerning  $\Gamma$ -convergence. Let  $(X, d)$  be a metric space, and let  $F, F_n : X \rightarrow \mathbb{R} \cup \{+\infty\}$  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  $\Gamma_X\text{-}\liminf_{n \rightarrow \infty} F_n(x)$  and  $\Gamma_X\text{-}\limsup_{n \rightarrow \infty} F_n(x)$ , respectively, are equal to  $F(x)$ , where

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

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 nonlocal Hessian.** The precise form we have chosen for the nonlocal Hessian can be derived from the model case of nonlocal gradients—the fractional gradient—which has been developed in [SS14]. Here we prove several results analogous to the first-order case, as in [MS15], for the generalizations involving generic radial weights that satisfy (1.4)–(1.6). Of primary importance is to first establish that the distributional nonlocal Hessian defined by (1.12) 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 (3.15) below together with the fact that  $\varphi$  is of compact support we have

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

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 what follows.

The second important item to address is the agreement of the distributional nonlocal Hessian with the nonlocal 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 that  $u \in L^p(\mathbb{R}^N)$  and  $\frac{|u(x+z) - 2u(x) + u(x-z)|^q}{|z|^{2q}} \rho_n(z) \in L^1(\mathbb{R}^N \times \mathbb{R}^N)$  for some  $1 \leq q \leq +\infty$ . 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

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

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

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

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

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

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 shown by alternating  $x$  and  $y$  and using the symmetry of the domain. Finally, (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 nonlocal Hessian and nonlocal 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 (1.11) exists whenever  $u \in C_c^2(\mathbb{R}^N)$ . However, this is also the case for functions in the spaces  $W^{2,p}(\mathbb{R}^N)$  and  $BV^2(\mathbb{R}^N)$ ; see Lemmas 3.1 and 3.4.

**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  $L = L(\epsilon, p) \gg 1$  such that

$$(3.4) \quad \sup_n \int_{B(0, LR)^c} |H_n v(x)|^p dx \leq \epsilon.$$

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 (3.4) 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 (3.4). 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

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

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

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

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 componentwise, i.e.,  $|(H_n u - \nabla^2 u)_{(i_0, j_0)}| \rightarrow 0$  by considering two cases  $i_0 \neq j_0$  and  $i_0 = j_0$ . Before we proceed, let us mention some useful facts. Observe first that Proposition 5.1 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

$$\begin{aligned} (3.5) \quad \int_{\mathbb{R}^N} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho_n(z) dz &= \int_0^\infty \rho_n(t) t^{N-1} dt \int_{\mathcal{S}^{N-1}} \nu_{i_0}^2 \nu_{j_0}^2 d\mathcal{H}^{N-1}(x) \\ &= \frac{1}{N(N+2)} \cdot \begin{cases} 1, & i_0 \neq j_0, \\ 3, & i_0 = j_0. \end{cases} \end{aligned}$$

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

$$(3.6) \quad \mathbf{i}_0 = \mathbf{j}_0 : \int_{\mathbb{R}^N} \frac{z_i z_{j_0}^3}{|z|^4} \rho_n(z) dz = 0 \quad \text{for } i \neq j_0,$$

$$(3.7) \quad \mathbf{i}_0 = \mathbf{j}_0 : \int_{\mathbb{R}^N} \frac{z_i z_j z_{j_0}^2}{|z|^4} \rho_n(z) dz = 0 \quad \text{for } i \neq j_0, j \neq j_0, i \neq j,$$

$$(3.8) \quad \mathbf{i}_0 \neq \mathbf{j}_0 : \int_{\mathbb{R}^N} \frac{z_i z_j z_{i_0} z_{j_0}}{|z|^4} \rho_n(z) dz = 0 \quad \text{for } i \neq i_0, j \neq j_0.$$

Subcase  $\mathbf{i}_0 \neq \mathbf{j}_0$ . Using (3.5), for the case that  $i_0 \neq j_0$ , we have

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

Moreover, (3.6)–(3.7) imply 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

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

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  $\delta > 0$ :

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

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

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

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

We therefore proceed to make estimates for (3.9). 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 (3.12) we can estimate

(3.10)

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

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

$$(3.12) \quad d^2 u(x,y) = (x-y)^T \left( \int_0^1 \int_0^1 \nabla^2 u(x + (t+s-1)(y-x)) ds dt \right) (x-y),$$

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  $\mathbf{i}_0 = \mathbf{j}_0$ .* Let us record several observations before we proceed with this case. In fact, the same argument shows that for a single  $i \in \{1, \dots, N\}$ ,

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

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

$$(3.14) \quad \left| \int_{\mathbb{R}^N} \frac{d^2 u(x,y) - (x-y)^T \nabla^2 u(x)(x-y)}{|x-y|^2} \rho_n(x-y) dy \right| \rightarrow 0.$$

Moreover, we observe that the same formula from Proposition 5.1 and cancellation of odd powers imply that

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

while we also have that

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

Thus, we can estimate

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

and the proof is completed by invoking the convergences established in (3.13) and (3.14). ■

**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 nonlocal Hessian  $H_n u$  converges to  $\nabla^2 u$  in  $L^p$ . The first step is to 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.

**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*

$$(3.15) \quad \int_{\mathbb{R}^N} |H_n u(x)|^p dx \leq M \|\nabla^2 u\|_{L^p(\mathbb{R}^N)}^p,$$

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

*Proof.* 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 nonlocal Hessian and utilizing Jensen's inequality, (3.12), and Fubini's theorem, we have the following successive estimates (the constant is always denoted with  $M(N, p)$ ):

$$\begin{aligned} (3.16) \quad &\int_{\mathbb{R}^N} |H_n v(x)|^p dx \\ &= \left( \frac{N(N+2)}{2} \right)^p \int_{\mathbb{R}^N} \left| \int_{\mathbb{R}^N} \frac{d^2 v(x, y)}{|x-y|^2} \frac{\left( (x-y) \otimes (x-y) - \frac{|x-y|^2}{N+2} I_N \right)}{|x-y|^2} \rho_n(x-y) dy \right|^p dx \\ &\leq M(N, p) \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{|d^2 v(x, y)|}{|x-y|^2} \rho_n(x-y) dy \right)^p dx, \end{aligned}$$



(3.17)

$$\begin{aligned}
&\leq M(N, p) \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{|d^2v(x, y)|^p}{|x - y|^{2p}} \rho_n(x - y) dy \right) dx \\
&\leq M(N, p) \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{\int_0^1 |\nabla v(x + t(y - x)) - \nabla v(x + (t - 1)(y - x))|^p dt}{|x - y|^p} \rho_n(x - y) dy \right) dx \\
&= M(N, p) \int_0^1 \int_0^1 \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} (|\nabla^2 v(x + (t + s - 1)(y - x))|^p) \rho_n(x - y) dy \right) dx ds dt \\
&= M(N, p) \int_0^1 \int_0^1 \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} (|\nabla^2 v(x + (t + s - 1)\xi)|^p) \rho_n(\xi) d\xi \right) dx ds dt, \\
&= M(N, p) \int_0^1 \int_0^1 \int_{\mathbb{R}^N} \rho_n(\xi) \|\nabla^2 v\|_{L^p(\mathbb{R}^N)}^p d\xi ds dt \\
&= M(N, p) \|\nabla^2 v\|_{L^p(\mathbb{R}^N)}^p.
\end{aligned}$$

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

$$(3.18) \quad \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{|v_k(x + z) - 2v_k(x) + v_k(x - z)|^p}{|z|^{2p}} \rho_n(z) dz \right) dx \leq M \|\nabla^2 v_k\|_{L^p(\mathbb{R}^N)}^p \quad \forall k \in \mathbb{N}.$$

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.

If we can establish that  $H_n u$  is well-defined as a Lebesgue integral, then Jensen's inequality and Fatou's lemma imply that

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

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 (3.17) holds for  $W^{2,p}$  functions as well, thus completing the proof.

Finally, the Gagliardo–Nirenberg inequality

$$\|\nabla^2 u\|_{L^1} \leq C \|\nabla^2 u\|_{L^p}^\theta \|u\|_{L^p}^{1-\theta}$$

implies that for  $u \in W^{2,p}$ ,  $\nabla^2 u \in L^1$ , which by the preceding display yields that  $H_n u$  is well-defined as a Lebesgue integral. ■

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 now use the fact 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

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

**3.3. Localization–BV<sup>2</sup>( $\mathbb{R}^N$ ) case.** Analogously with the first-order case in [MS15], we can define a second-order nonlocal divergence that corresponds to  $H_n$ , and we can also derive a second-order nonlocal integration by parts formula which is an essential tool for the proofs of this section. The second-order nonlocal divergence is defined for a function  $\phi = (\phi_{ij})_{i,j=1}^N$  as

(3.19)

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

where  $A \cdot B = \sum_{i,j=1}^N A_{ij} B_{ij}$  for two  $N \times N$  matrices  $A$  and  $B$ . Notice that (3.19) is well-defined for  $\phi \in C_c^2(\mathbb{R}^N, \mathbb{R}^{N \times N})$ .

**Theorem 3.2 (second-order nonlocal integration by parts formula).** *Suppose that  $u \in L^1(\mathbb{R}^N)$  and  $\frac{|d^2 u(x,y)|}{|x-y|^2} \rho_n(x-y) \in L^1(\mathbb{R}^N \times \mathbb{R}^N)$ , and let  $\phi \in C_c^2(\mathbb{R}^N, \mathbb{R}^{N \times N})$ . Then*

$$(3.20) \quad \int_{\mathbb{R}^N} H_n u(x) \cdot \phi(x) dx = \int_{\mathbb{R}^N} u(x) \mathcal{D}_n^2 \phi(x) dx.$$

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 nonlocal divergence to the continuous analogue  $\operatorname{div}^2 \phi$ , where  $\phi \in C_c^2(\mathbb{R}^N, \mathbb{R}^{N \times N})$  and

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

**Lemma 3.3.** Let  $\phi \in C_c^2(\mathbb{R}^N, \mathbb{R}^{N \times N})$ . Then for every  $1 \leq p \leq \infty$  we have

$$(3.21) \quad \lim_{n \rightarrow \infty} \|\mathcal{D}_n^2 \phi - \operatorname{div}^2 \phi\|_{L^p(\mathbb{R}^N)} = 0.$$

*Proof.* The proof follows immediately from Theorem 1.4.  $\blacksquare$

The following lemma shows that the nonlocal Hessian (1.11) 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.

**Lemma 3.4.** Suppose that  $u \in \operatorname{BV}^2(\mathbb{R}^N)$ . Then  $H_n u \in L^1(\mathbb{R}^N, \mathbb{R}^{N \times N})$  with

$$(3.22) \quad \int_{\mathbb{R}^N} |H_n u(x)| dx \leq M |D^2 u|(\mathbb{R}^N),$$

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

*Proof.* 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 way similar to how it was used in Lemma 3.1, we can establish that  $H_n u$  is well-defined as a Lebesgue integral, along with the estimate

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

where above we employed the strict convergence of  $D^2 u_k$  to  $D^2 u$ . Thus the result has been demonstrated.  $\blacksquare$

We can now proceed to prove the localization result for  $\operatorname{BV}^2$  functions. Recall that we defined  $\mu_n$  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.* We first proceed to prove (1.14) for  $C_c^\infty$  functions, and then we conclude with a density argument. From the estimate (3.22) we have that  $(|\mu_n|)_{n \in \mathbb{N}}$  is bounded; thus there exist a subsequence  $(\mu_{n_k})_{k \in \mathbb{N}}$  and an  $\mathbb{R}^{N \times N}$ -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 (3.20) and Lemma 3.3 we get that

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

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

Now we let  $\phi \in C_0(\mathbb{R}^N, \mathbb{R}^{N \times N})$  and  $\epsilon > 0$ . From the density of  $C_c^\infty(\mathbb{R}^N, \mathbb{R}^{N \times N})$  in  $C_0(\mathbb{R}^N, \mathbb{R}^{N \times N})$  we can find a  $\psi \in C_c^\infty(\mathbb{R}^N, \mathbb{R}^{N \times N})$  such that  $\|\phi - \psi\|_\infty < \epsilon$ . Then, using also the estimate (3.22), we have

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

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

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

and since  $\epsilon$  is arbitrary, we have (1.14).  $\blacksquare$

Let us note here that in the case  $N = 1$  we can also prove strict convergence of the measures  $\mu_n$  to  $D^2u$ ; that is, in addition to (1.14) we also have

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

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

$$(3.23) \quad \mu_n \rightarrow D^2u \text{ weakly}^*, \text{ and}$$

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

*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

$$(3.25) \quad |D^2u|(\mathbb{R}) \leq \liminf_{n \rightarrow \infty} |\mu_n|(\mathbb{R}).$$

Thus it suffices to show that

$$(3.26) \quad \limsup_{n \rightarrow \infty} |\mu_n|(\mathbb{R}) \leq |D^2u|(\mathbb{R}).$$

Note that in dimension one the nonlocal Hessian formula is

$$(3.27) \quad H_n u(x) = \int_{\mathbb{R}} \frac{u(y) - 2u(x) + u(x + (x - y))}{|x - y|^2} \rho_n(x - y) dy.$$

Following the proof of Lemma 3.1, we can easily verify that for  $v \in C^\infty(\mathbb{R}) \cap BV^2(\mathbb{R})$  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 (3.15) is equal to 1. Using Fatou's lemma and the  $BV^2$  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 (3.26) straightforwardly follows.  $\blacksquare$

**3.4. Characterization of higher-order Sobolev and BV spaces.** Characterization of Sobolev and BV spaces in terms of nonlocal, derivative-free energies has been done so far only in the first-order case; see [BBM01, Pon04b, Men12, MS15]. Here we characterize the spaces  $W^{2,p}(\mathbb{R}^N)$  and  $BV^2(\mathbb{R}^N)$  using our definition of nonlocal Hessian.

*Proof of Theorem 1.7.* First, we prove (1.15). 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, the sequence  $\mathfrak{H}_n u$  is representable (up to a subsequence) by a sequence of functions 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  $\psi \in C_c^\infty(\mathbb{R}^N)$ ,

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

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, eq. 2.5])

$$(3.28) \quad \|\nabla u\|_{L^p(\mathbb{R}^N)} \leq C \|\nabla^2 u\|_{L^p(\mathbb{R}^N)}^{\frac{1}{2}} \|u\|_{L^p(\mathbb{R}^N)}^{\frac{1}{2}}$$

implies that the first distributional derivative 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 (1.16). Again supposing that  $u \in \text{BV}^2(\mathbb{R}^N)$  we have that Lemma 3.4 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 exist 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,

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

something that shows that  $\mu = D^2 u$ . Again, by first mollifying and then passing the limit, the inequality (3.28) implies that  $Du \in \mathcal{M}(\mathbb{R}^N, \mathbb{R}^N)$ . However,  $Du \in \mathcal{M}(\mathbb{R}^N, \mathbb{R}^N)$  and  $D^2 u \in \mathcal{M}(\mathbb{R}^N, \mathbb{R}^{N \times N})$  imply that  $Du$  is an  $L^1(\mathbb{R}^N, \mathbb{R}^N)$  function (which is a simple consequence of the Sobolev inequality; but see also [AFP00, Exerc. 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

$$(3.29) \quad F_n(u) := \begin{cases} \int_{\mathbb{R}^N} |\mathfrak{H}_n u| dx & \text{if } \mathfrak{H}_n u \text{ is representable by an } L^1 \text{ function,} \\ \infty & \text{otherwise.} \end{cases}$$

*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 nonlocal Hessian and the convergence  $u_n \rightarrow u$  in  $L^1(\mathbb{R}^N)$  imply that

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

We thus conclude that

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

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

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

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 that

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

Now, taking the lower semicontinuous envelope with respect to  $L^1(\mathbb{R}^N)$  strong convergence, using that both the  $\Gamma_{L^1(\mathbb{R}^N)}$ -lim sup and the mapping  $u \mapsto |D^2u|(\mathbb{R}^N)$  are lower semicontinuous on  $L^1(\mathbb{R}^N)$  (for the  $\Gamma$ -lim sup see [DM93, Prop. 6.8]), we deduce that

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

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 nonlocal definition of  $H_n$  as in (1.11) localizes to the classical distributional Hessian for a specific choice of the weights  $\rho_n$  and thus can be rightfully called a nonlocal Hessian. In numerical applications, however, the strength of such nonlocal models lies in the fact that the weights can be chosen to have nonlocal interactions and model specific patterns in the data. A classic example is the nonlocal total variation [GO08]:

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

A possible choice is to choose  $w(x,y)$  large if the *patches* (neighborhoods) 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 [GO08] this is achieved by setting  $w(x, y) = 1$  if the neighborhood around  $y$  is one of the  $K \in \mathbb{N}$  closest to the neighborhood 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 (1.11) 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 nonlocal generalizations of functionals such as  $\text{TV}^2$ , for example to minimize the nonlocal  $L^2$ - $\text{TV}^2$  model

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

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 symmetrized 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 (1.11) that allows for full nonsymmetric weights. We start with the realization that the finite-difference 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 nonlocal gradient  $G'_u(x)$  and Hessian looking for a nonlocal 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 (1.11). We denote definition (4.4) the *implicit* nonlocal 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 us to incorporate nonlocal 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 nonlocal 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 precomputation step; however, we did not find this to increase overall performance.

While the implicit and explicit models 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 theorem.

**Theorem 4.1.** *Let  $\rho$  be a radially symmetric function satisfying the conditions (1.4)–(1.6), let  $u \in \text{BV}^2(\mathbb{R}^N)$ , with  $\nabla u$  being Lipschitz, let  $x \in \mathbb{R}^N$ , and let*

$$(4.5) \quad (G'_u(x), H'_u(x)) = \underset{G_u \in \mathbb{R}^N, H_u \in \text{Sym}(\mathbb{R}^{N \times N})}{\text{argmin}} \frac{1}{2} \int_{\mathbb{R}^N} \left( u(x+z) - u(x) - G^\top z - \frac{1}{2} z^\top H z \right)^2 \frac{\rho(z)}{|z|^4} dz.$$

Then the optimal  $H'_u(x)$  is given by the explicit nonlocal Hessian, i.e.,

$$H'_u(x) = H_u(x) = \frac{N(N+2)}{2} \int_{\mathbb{R}^N} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \frac{z \otimes z - \frac{|z|^2}{N+2} I_N}{|z|^2} \rho(z) dz.$$

This means that by setting  $\Omega = \mathbb{R}^N$ , assuming some extra regularity for  $u$ , and substituting the weights  $\sigma_x(z)$  with  $\rho(z)/|z|^4$  the implicit nonlocal Hessian coincides with the explicit one. In order to prove Theorem 4.1 we need first the following lemma, whose only difference is that we are integrating over  $\mathbb{R}^N \setminus B(0, \epsilon)$  in order to deal with the introduced singularity  $1/|z|^4$  at the origin.

**Lemma 4.2.** *Let  $\rho$  be a radially symmetric function satisfying the conditions (1.4)–(1.6), let  $u \in \text{BV}^2(\mathbb{R}^N)$ , let  $x \in \mathbb{R}^N$ , and let*

$$(4.6) \quad (G', H') = \underset{G \in \mathbb{R}^N, H \in \text{Sym}(\mathbb{R}^{N \times N})}{\text{argmin}} \frac{1}{2} \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \left( u(x+z) - u(x) - G^\top z - \frac{1}{2} z^\top H z \right)^2 \frac{\rho(z)}{|z|^4} dz.$$

Then the optimal  $H'$  is given by

$$H'_{ij} = C_\epsilon \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \frac{z_i z_j}{|z|^2} \rho(z) dz, \quad i \neq j,$$

where  $C_\epsilon$  is a constant depending only on  $\epsilon$ ,  $\rho$ ,  $N$  and satisfying  $C_\epsilon \xrightarrow{\epsilon \rightarrow 0} N(N+2)/2$ . For the case  $i = j$ ,

$$(H'_{ii})_{i=1, \dots, N} = D_\epsilon \cdot \left( \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \frac{z_{i_0}^2}{|z|^2} \rho(z) dz \right)_{i_0=1, \dots, N}$$

holds, where  $D_\epsilon \xrightarrow{\epsilon \rightarrow 0} \frac{N(N+2)}{2} (I - \frac{1}{N+2} E)$  and  $E$  is the all-ones matrix. Here “ $\cdot$ ” denotes matrix-vector multiplication.

*Proof.* The optimality conditions for (4.6) in terms of  $G$  and  $H$  are

$$(4.7) \quad \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \left( u(x+z) - u(x) - G^\top z - \frac{1}{2} z^\top H z \right) z \frac{\rho(z)}{|z|^4} dz = 0,$$

$$(4.8) \quad \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \left( u(x+z) - u(x) - G^\top z - \frac{1}{2} z^\top H z \right) z z^\top \frac{\rho(z)}{|z|^4} dz = 0$$

under the constraint  $H \in \text{Sym}(\mathbb{R}^{N \times N})$ . Note that from a nonsymmetric solution  $H_0$  we can always construct a symmetric solution  $H \in \text{Sym}(\mathbb{R}^{N \times N})$  by letting  $H = \frac{1}{2}(H_0 + H_0^\top)$ , as the equations are invariant under transposition of  $H$ . The above conditions correspond to the following sets of  $N$  and  $N^2$  equations, respectively:

$$\begin{aligned} & \int_{\mathbb{R}^N \setminus B(0, \epsilon)} z_i (u(x+z) - u(x)) \frac{\rho(z)}{|z|^4} dz \\ &= \left\langle \int_{\mathbb{R}^N \setminus B(0, \epsilon)} z_i z \frac{\rho(z)}{|z|^4} dz, G \right\rangle + \left\langle \int_{\mathbb{R}^N \setminus B(0, \epsilon)} z_i \left( \frac{1}{2} z z^\top \right) \frac{\rho(z)}{|z|^4} dz, H \right\rangle, \quad i = 1, \dots, N, \end{aligned}$$

and

$$\begin{aligned} & \int_{\mathbb{R}^N \setminus B(0, \epsilon)} z_i z_j (u(x+z) - u(x)) \frac{\rho(z)}{|z|^4} dz \\ &= \left\langle \int_{\mathbb{R}^N \setminus B(0, \epsilon)} z_i z_j z \frac{\rho(z)}{|z|^4} dz, G \right\rangle + \left\langle \int_{\mathbb{R}^N \setminus B(0, \epsilon)} z_i z_j \left( \frac{1}{2} z z^\top \right) \frac{\rho(z)}{|z|^4} dz, H \right\rangle, \quad i, j = 1, \dots, N. \end{aligned}$$

Note that  $G$  and  $H$  are elements of  $\mathbb{R}^N$  and  $\mathbb{R}^{N \times N}$ , and both sides of the inner products are finite-dimensional vectors (respectively, matrices).

If we collect the entries of  $G$  and  $F$  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 submatrices  $V$  are all of the form

$$\int_{\mathbb{R}^N \setminus B(0, \epsilon)} z_i z_j z_k \frac{\rho(z)}{|z|^4} dz, \quad i, j, k \in \{1, \dots, N\}.$$

No matter what the choice of  $i, j, k$  is, there is always one index with an odd power, so every one of these integrals is zero due to symmetry and the fact that  $\rho$  is even, i.e.,  $\rho(z) = \rho(-z)$ . This means that the conditions on the gradient and the Hessian parts of  $p$  decouple, i.e., the problem is

$$(4.9) \quad A p_G = a, \quad B p_H = b.$$

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

$$\frac{1}{2} \int_{\mathbb{R}^N \setminus B(0, \epsilon)} z_i z_j z_{i_0} z_{j_0} \frac{\rho(z)}{|z|^4} dz.$$

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

$$\frac{1}{2} \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{z_k^4}{|z|^4} \rho(z) dz \quad \text{for some } k$$

or

$$\frac{1}{2} \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{z_k^2 z_l^2}{|z|^4} \rho(z) dz \quad \text{for some } k \neq l.$$

The right-hand side vector  $b$  in (4.9) is

$$\begin{aligned} b_{ij} &= \int_{\mathbb{R}^N \setminus B(0, \epsilon)} (u(x+z) - u(x)) \frac{z_i z_j}{|z|^4} \rho(z) dz \\ &= \frac{1}{2} \int_{\mathbb{R}^N \setminus B(0, \epsilon)} (u(x+z) - 2u(x) + u(x-z)) \frac{z_i z_j}{|z|^4} \rho(z) dz. \end{aligned}$$

Thus in the end we obtain the following  $N^2$  equations for  $p_H = (p_{ij})_{i,j=1 \dots N}$ :

$$(4.10) \quad \sum_{i,j} p_{ij} \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{z_i z_j z_{i_0} z_{j_0}}{|z|^4} \rho(z) dz = \int_{\mathbb{R}^N \setminus B(0, \epsilon)} (u(x+z) - 2u(x) + u(x-z)) \frac{z_{i_0} z_{j_0}}{|z|^4} \rho(z) dz,$$

where  $i_0, j_0 = 1, \dots, N$ .

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

*Case  $i_0 \neq j_0$ .* All the terms in the left-hand side of (4.10) vanish except for  $(i, j) = (i_0, j_0)$  and  $(i, j) = (j_0, i_0)$ . Since we also require  $H = p_{ij}$  to be symmetric, we get the condition

$$2p_{i_0 j_0} \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho(z) dz = \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{u(x+z) - 2u(x) + u(x-z)}{|z|^2} \frac{z_{i_0} z_{j_0}}{|z|^2} \rho(z) dz.$$

Then we are done for this case by simply observing that

$$C_\epsilon^{-1} := 2 \int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho(z) dz \xrightarrow{\epsilon \rightarrow 0} 2 \int_{\mathbb{R}^N} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho(z) dz = \frac{2}{N(N+2)},$$

where the last equality follows from (3.5).

*Case  $i_0 = j_0$ .* All the terms in the left-hand side of (4.10) vanish apart for  $i = j$ ; i.e., we have

$$\sum_i p_{ii} \underbrace{\int_{\mathbb{R}^N \setminus B(0, \epsilon)} \frac{z_i^2 z_{i_0}^2}{|z|^4} \rho(z) dz}_{M_\epsilon^{i, i_0}} = \int_{\mathbb{R}^N \setminus B(0, \epsilon)} (u(x+z) - 2u(x) + u(x-z)) \frac{z_{i_0}^2}{|z|^4} \rho(z) dz,$$

where  $i_0 = 1, \dots, N$ . Again from (3.5) it follows that

$$M_\epsilon \xrightarrow{\epsilon \rightarrow 0} \frac{1}{N(N+2)} (E + 2I),$$

where  $E$  is the all-ones matrix. Thus

$$D_\epsilon := M_\epsilon^{-1} \xrightarrow{\epsilon \rightarrow 0} \frac{N(N+2)}{2} \left( I - \frac{1}{N+2} E \right),$$

and the lemma has been shown.  $\blacksquare$

*Remark 4.3.* Note that we did not have to explicitly compute the optimal gradient  $G$  in the lemma above. However, we can bound  $G$  uniformly in  $\epsilon$ . For that we assume that  $\nabla u$  is Lipschitz, say with constant  $L$ . Indeed, the gradient part in (4.9) reads as

$$G_i \int_{\mathbb{R}^N \setminus B(0,\epsilon)} \frac{z_i^2}{|z|^4} \rho(z) dz = \int_{\mathbb{R}^N \setminus B(0,\epsilon)} (u(x+z) - u(x)) \frac{z_i}{|z|^4} \rho(z) dz, \quad i = 1, \dots, N.$$

By symmetry we have

$$\int_{\mathbb{R}^N} \frac{z_i^2}{|z|^4} \rho_n(z) dz = \frac{1}{N} \int_{\mathbb{R}^N} \frac{1}{|z|^2} \rho_n(z) dz;$$

therefore from the first equation we conclude the bound

$$\begin{aligned} |G_i| \int_{\mathbb{R}^N \setminus B(0,\epsilon)} \frac{\rho(z)}{|z|^2} dz &= N \left| \int_{\mathbb{R}^N \setminus B(0,\epsilon)} \frac{u(x+z) - u(x)}{|z|} \frac{z_i}{|z|} \frac{\rho(z)}{|z|^2} dz \right| \\ &\leq LN \int_{\mathbb{R}^N \setminus B(0,\epsilon)} \frac{\rho(z)}{|z|^2} dz \\ &\Rightarrow |G_i| \leq LN. \end{aligned}$$

We are now ready to prove Theorem 4.1, which is based on Lemma 4.2, Remark 4.3, and a  $\Gamma$ -convergence argument.

*Proof of Theorem 4.1.* Let  $F$  and  $F_\epsilon$  be the functions mapping a point  $x$  to the minimizers of (4.5) and (4.6). From Fatou's lemma the functionals  $F_\epsilon$  are lower semicontinuous. Also

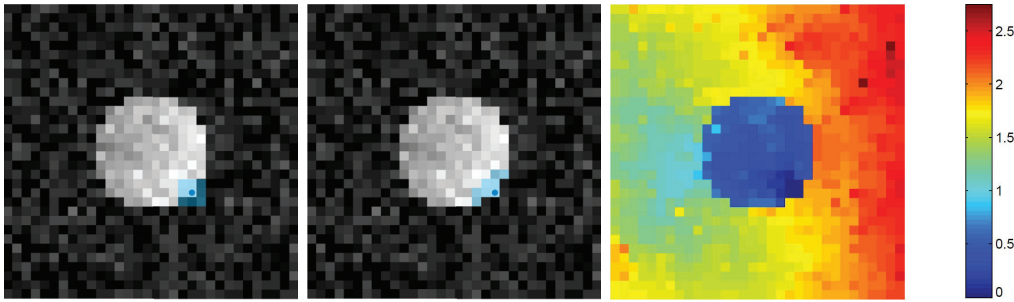
$$F_\epsilon \rightarrow F, \quad \text{increasingly, pointwise,}$$

and thus from [DM93, Rem. 5.5] we have that the functionals  $F_\epsilon$   $\Gamma$ -converge to  $F$  as  $\epsilon \rightarrow 0$ . Finally, from the fact that  $u \in \text{BV}^2(\mathbb{R}^N)$  and Remark 4.3 the minimizers of  $F_\epsilon$  are bounded uniformly in  $\epsilon$ . Then [DM93, Thm. 7.4], we have that  $F$  attains its minimum and

$$\operatorname{argmin} F = \lim_{\epsilon \rightarrow 0} \operatorname{argmin} F_\epsilon = Hu(x),$$

where the  $\operatorname{argmin}$  here refers to the Hessian part of the minimizers.  $\blacksquare$

Note that while Theorem 4.1 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 (4.5).



**Figure 2.** Adaptive choice of the neighborhood 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  $3 \times 3$  neighborhood. Center: With a suitable choice of the weights  $\sigma_x$ , the discretization of the (nonlocal) Hessian at the marked point only (or mostly) involves points that are likely to belong to the same affine region, here the inside of the disc. This allows one to use a straightforward 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 neighborhood of  $x$  that is used to define the nonlocal Hessian at  $x$ .

**4.2. Choosing the weights for jump preservation.** A characteristic of nonlocal 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 to overcome this issue, most notably approaches based on combined first- and higher-order functionals (see [PS14] and the references therein), infimal convolution [CL97], and TGV [BKP10, SST11]. Here we propose another strategy, making use of the nonlocal formulation (4.4). Note that for general fixed weights  $\rho$ , even in the implicit model, finiteness of  $\int_{\Omega} |H_{\rho}| dx$  does not require existence of  $|D^2u|$ , as Theorem 1.7 only concerns (specific) sequences of  $\rho$ .

We draw our motivation for choosing the weights partly from a recent discussion of nonlocal “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 extending this idea to the higher-order energy minimization framework (Figure 2).

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 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

$$(4.11) \quad \varphi(\nabla g) := |\nabla g|^2 + \gamma$$

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  $d_{\mathcal{M}}(x, y)$  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 neighbors  $y^1, y^2, \dots$  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 neighborhood size  $M \in \mathbb{N}$  and set the weights to

$$(4.12) \quad \sigma_x(y) := \begin{cases} 1, & i \leq M, \\ 0, & i > M. \end{cases}$$

In other words, the nonlocal Hessian at  $x$  is computed using its  $M$  closest neighbors 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

$$(4.13) \quad |\nabla c(y)| = \varphi(\nabla g(y)),$$

$$(4.14) \quad c(x) = 0$$

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 neighbors 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 nonlocal Hessian in (4.4) becomes underdetermined. In our experiments we found  $M = 12$  to be a good compromise, but the choice does not appear to be a very critical one. In the experiments we used the  $L^1$ -nonlocal TV<sup>2</sup> model

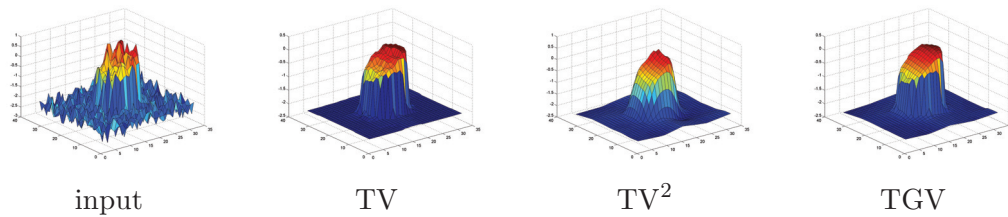
$$(4.15) \quad \min_{\substack{u: \Omega \rightarrow \mathbb{R}, \\ G'_u: \Omega \rightarrow \mathbb{R}^N, \\ H'_u: \Omega \rightarrow \text{Sym}(\mathbb{R}^N \times \mathbb{R}^N)}} \sum_{x \in \Omega} |u(x) - g(x)|^p dx + \alpha \sum_{x \in \Omega} \omega(x) |H'_u(x)| dx$$

$$(4.16) \quad \text{s.t.} \quad A \begin{pmatrix} G'_u \\ H'_u \end{pmatrix} = Bu,$$

where  $\alpha > 0$  is the regularization strength, and  $p \in \{1, 2\}$ . The linear constraints implement the optimality conditions for  $G'_u$  and  $H'_u$  from (4.4), similar to (4.7)–(4.8):

$$(4.17) \quad \sum_{y \in \Omega, z=y-x} (u(y) - u(x) - G'_u(x)^\top z - \frac{1}{2} z^\top H'_u(x) z) z \sigma_x(z) = 0, \quad x \in \Omega,$$

$$(4.18) \quad \sum_{y \in \Omega, z=y-x} (u(y) - u(x) - G'_u(x)^\top z - \frac{1}{2} z^\top H'_u(x) z) z z^\top \sigma_x(z) = 0, \quad x \in \Omega,$$



**Figure 3.** 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 an  $L^1$ -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 oversmoothing jumps. Total generalized variation (TGV,  $\alpha_0 = 1.5, \alpha_1 = 1.1$ ) performs best but still clips the slope at the top.

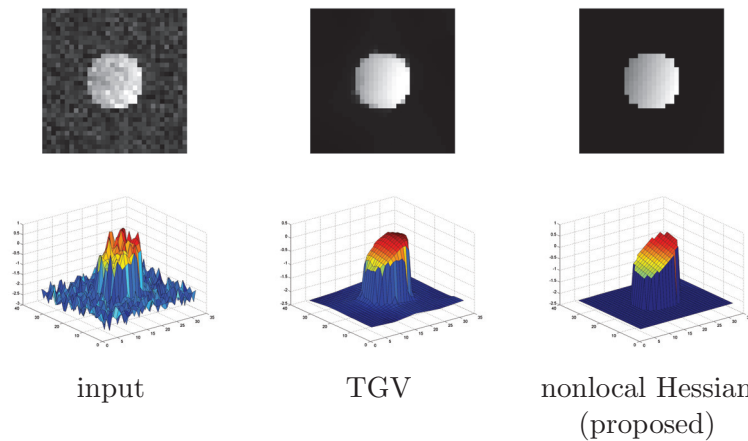
with a suitable parametrization of  $H'_u$  to enforce symmetry. Note that unlike the radially symmetric case in Lemma 4.2, the conditions do not necessarily uncouple; therefore the first-order component  $G'_u$  needs to be included in the optimization as well.

The local weight  $\omega$  is set as  $\omega(x) = M/|\{y \in \Omega | B_{y,x} \neq \emptyset\}|$ . 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 neighbors 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 in which a particular value  $u(x)$  appears.

**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, and the 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 the system matrix use a custom C++ implementation. For solving the minimization problems we used the commercial interior-point based Mosek solver with the CVX interface. This allows us to efficiently compute solutions with very high accuracy and therefore to evaluate the model without the risk of accidentally comparing only approximate solutions. The stopping criterium was set to guarantee that the normalized difference between the objective values  $f'$  of the numerical solution and  $f^*$  of the true minimizer satisfies  $(f' - f^*)/f^* \leq \sqrt{\epsilon} = 1.5 \cdot 10^{-8}$ . Alternatively, nonsmooth first-order methods could be used; however, in our experience they become prohibitively slow for a precision beyond  $10^{-4}$  due to the sublinear convergence rate.

Figure 3 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 4.** Nonlocal regularization of the problem in Figure 3. The adaptive choice of the neighborhood and weights together with the nonlocal Hessian preserves the jumps, clips the top of the slope, and allows one to perfectly reconstruct the piecewise affine signal.

Figure 4 shows the effect of our nonlocal 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.

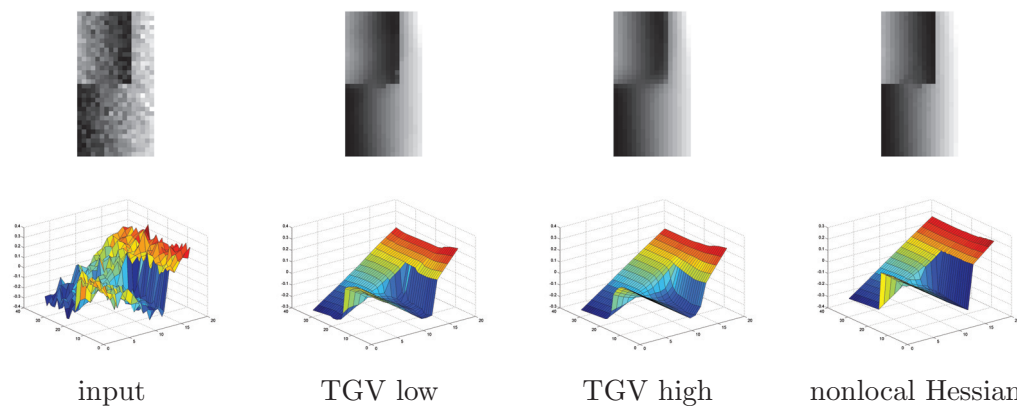
To see what happens with smaller jumps, we generated a pattern of opposing slopes (Figure 5). As expected, both TGV and the nonlocal Hessian approach struggle when the jump is small. This shows the limitations of our approach for choosing the weights—while it adds some structural information to the regularizer, this information is still restricted to a certain neighborhood of each point and does not take into account the full global structure.

Figures 6–8 show a quantitative comparison of our nonlocal 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 nonlocal 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 nonlocal 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 6), which is in line with the good results obtained with TGV for natural 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 the reader to [LBL13] for a discussion of a problem that is more demanding in terms of the correct choice of the regularization.

Finally, Figure 9 shows an application to the “cameraman” image with  $L^2$  data term. Small details on the camera as well as long, thin structures such as 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





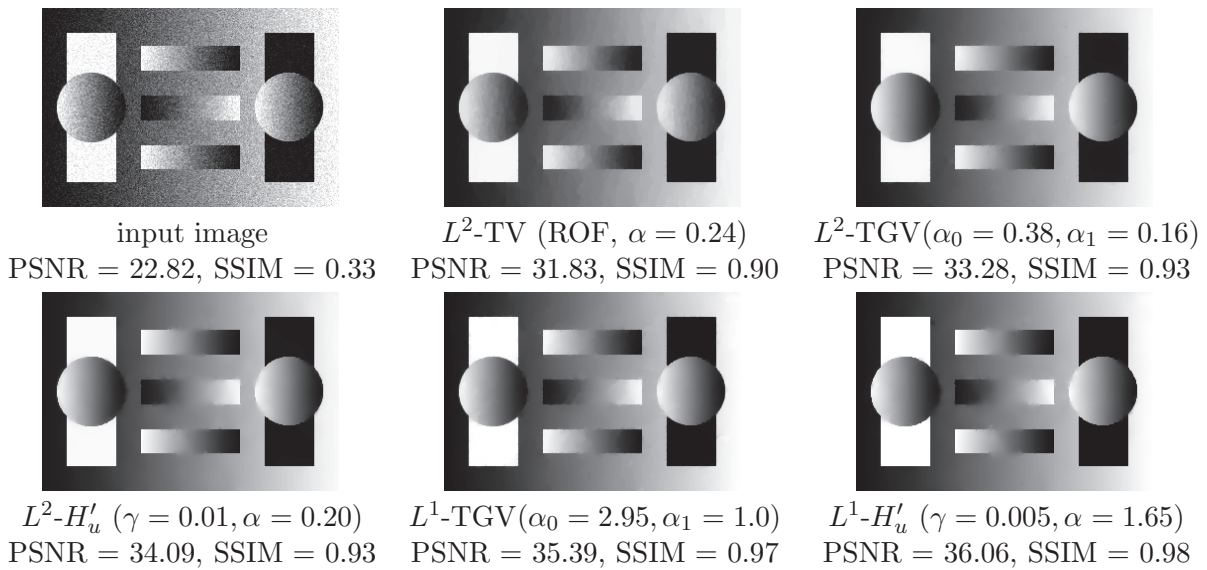
**Figure 5.** Denoising results for the “opposing slopes” image. The small jump at the crossing causes a slight amount of smoothing for both the TGV and the nonlocal approaches. TGV with low regularization strength (TGV low,  $\alpha_0 = \alpha_1 = 0.8$ ) does reasonably well at preserving the strong 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 nonlocal 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.

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, resulting in a further reduction of noise (bottom right image in Figure 9).

A current limitation is that our choice of weights can result in a pixelated structure along slanted edges, as seen in Figures 10 and 11. 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 neighboring points  $x, y$  the weight  $\sigma_x(y)$  is always at least some positive constant; however, we have not pursued this direction further. Whether this is a desired behavior or not depends on the underlying data—for applications such as segmentation a pixelwise decision might actually be preferable.

**5. Conclusion.** From the perspective of the analysis, the study of the nonlocal Hessian is a natural extension of the study of nonlocal gradients. While one has the straightforward observation we mention in the introduction—that one may use nonlocal gradients to characterize higher-order Sobolev spaces—the results of this paper alongside those of [MS15] provide a framework for general characterizations of Sobolev spaces of arbitrary (nonnegative integer) order. This notion, that one can write down an integrated Taylor series and use this as a definition of a nonlocal 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 Ponce [Pon04a], the lack of monotonicity of the nonlocal gradient has proven difficult in adapting



**Figure 6.** Denoising results for a geometric test image using the total variation (TV), total generalized variation (TGV), and nonlocal Hessian (NL-H) models (see Figure 7 for the discussion). For the nonlocal models  $M = 12$  neighbors were used.

his argument to our setting. This motivates the following open question.

*Open question.* Can one find a 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

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

for all  $n$  sufficiently large?

In fact, beyond the multitude of questions one could explore by making a comparison of the results for nonlocal gradients and Hessians with known results in the Sobolev spaces, there are already several interesting questions for nonlocal 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

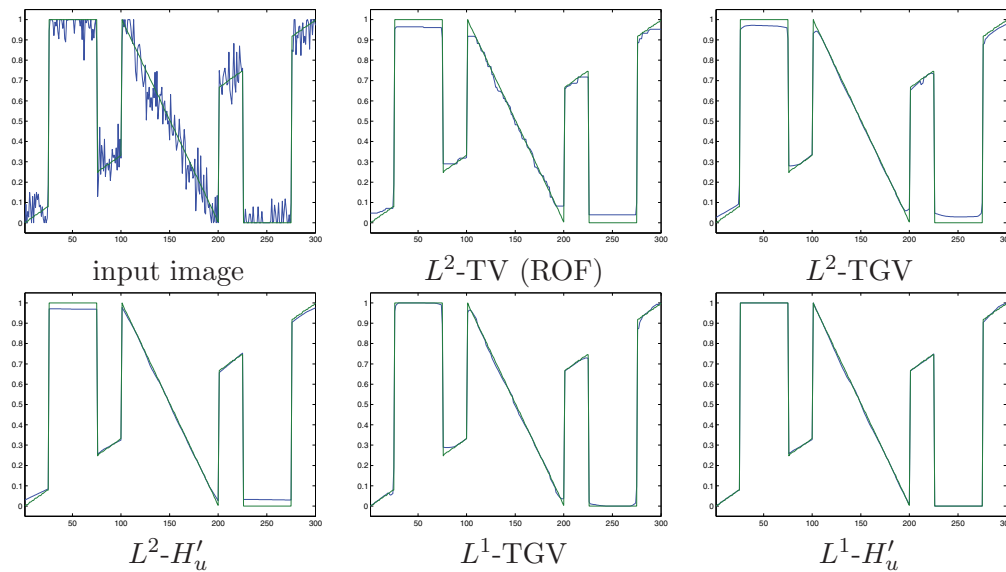
$$\int_{\Omega} |G_n u| dx \rightarrow |Du|(\Omega).$$

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 nonlocal gradient and the nonlocal Hessian when  $p = 1$ . In both the



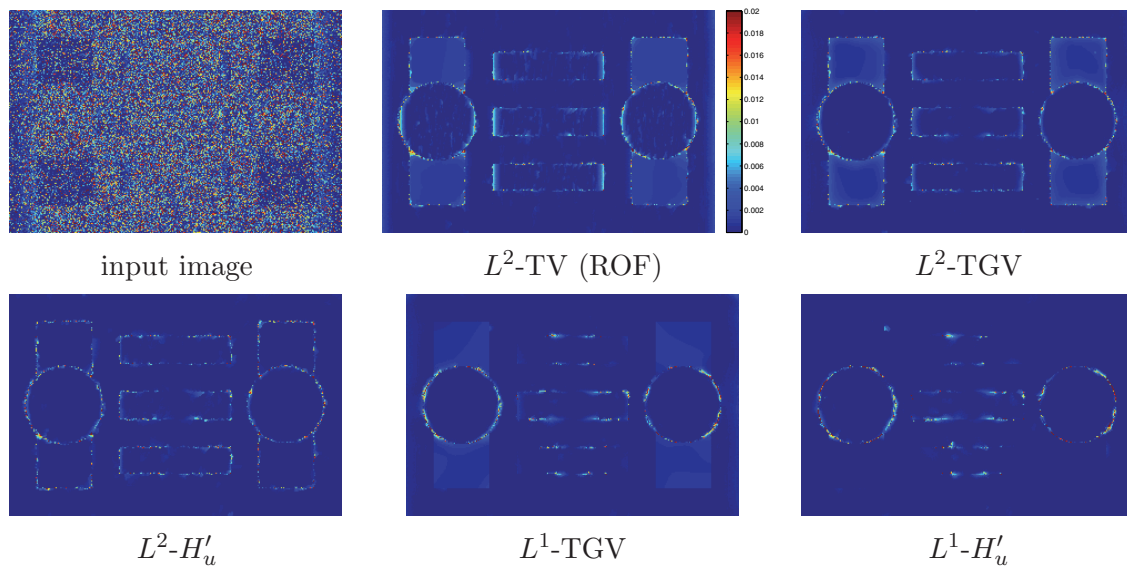
**Figure 7.** Slice at row 50 through the results images in Figure 6. The  $L^2$ -TV model shows staircasing as expected.  $L^2$ -TGV smooths out sharp local extrema, which is preserved better by the nonlocal  $L^2$  model. Changing the data term to  $L^1$ , as in the nonlocal  $L^1$  model, additionally removes the contrast reduction, even though the original noisy data is clipped to the interval  $[0, 1]$ .

paper [MS15] 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 nonlocal gradient or nonlocal 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 nonlocal gradient or nonlocal 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 nonlocal 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 already yields 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 nonlocal Hessian—jump-preserving second-order regularization—the approach is still somewhat local in the sense that we still restrict ourselves to a neighborhood of each point. It would be interesting to find an application that allows us 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 nonlocal Hessian (4.4) by a cubic- or even higher-order model.



**Figure 8.** Distribution of the  $L^2$  error to the ground truth for the experiment in Figure 6. In the TGV case the errors spread out from the jumps. This is greatly reduced when using the nonlocal Hessian, at the cost of introducing a few more local errors.

### Appendix.

**Proposition 5.1.** For  $N \geq 1$  we have the following definition and identity for  $C$ :

$$(5.1) \quad C_{i_0 j_0} := \int_{S^{N-1}} \nu_{i_0}^2 \nu_{j_0}^2 d\mathcal{H}^{N-1}(x) = \frac{|S^{N-1}|}{N(N+2)} \cdot \begin{cases} 1, & i_0 \neq j_0, \\ 3, & i_0 = j_0. \end{cases}$$

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

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

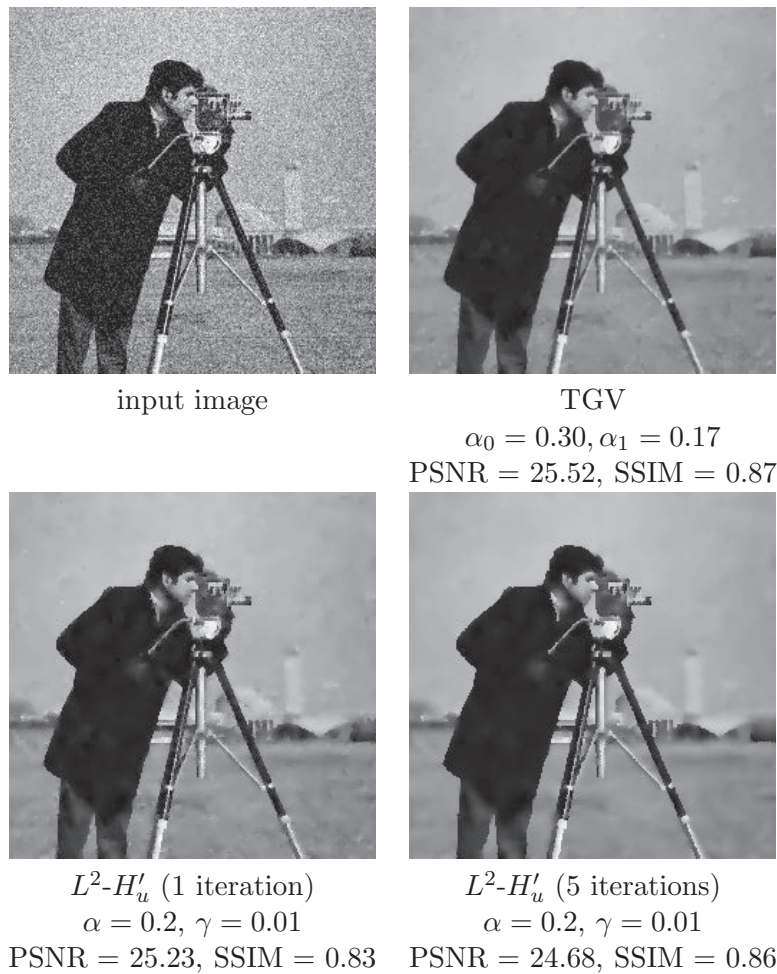
where  $E \in \mathbb{R}^{N \times N}$  is the all-ones matrix.

*Proof.* 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

$$(5.2) \quad \int_0^{2\pi} \cos^2 \alpha \sin^2 \alpha d\alpha = \frac{\pi}{4} \quad \text{or} \quad \int_0^{2\pi} \cos^2 \alpha \cos^2 \alpha d\alpha = \frac{3\pi}{4},$$

which agrees with (5.1). For the general case  $N \geq 3$ , we apply a general form of the coarea formula in [AFP00, eq. (2.72)] using  $E = S^{N-1}$ ,  $M = N$ ,  $N = N - 1$ ,  $k = 2$ ,  $f(x_1, x_2, \dots) = (x_1, x_2)$ ; then

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

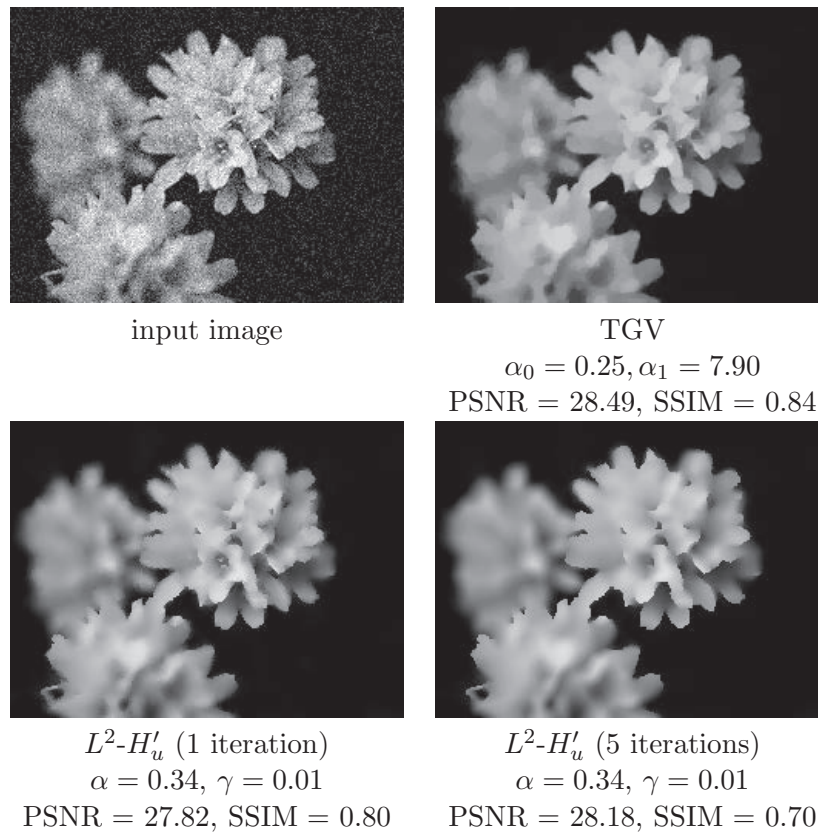


**Figure 9.** Nonlocal Hessian-based denoising with  $L^2$  data fidelity ( $p = 2$ ) on a real-world noisy input image (top left) with  $\sigma = 0.1$  Gaussian noise. Using a suitable choice of weights with a nonparametric neighborhood, small details and long, thin structures such as the camera handle 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. This effect can also be observed on other real-world images (Figures 10 and 11).

Here  $d^E f_x$  denotes the tangential differential of the function  $f$  at the point  $x$ , i.e., a linear map from the  $(N - 1)$ -dimensional tangent space  $\text{Tan}(E, x) = \text{Tan}(\mathcal{S}^{N-1}, x)$  at the point  $x \in \mathcal{S}^{N-1}$  into  $\mathbb{R}^2$  [AFP00, Def. 2.89], and  $C_k L$  denotes the coarea factor  $\sqrt{\det LL^\top}$  for a linear map  $L$ . In our case we set  $g(x) = cx_1^2 x_2^2$  with  $c = (1 - x_1^2 - x_2^2)^{-1/2}$ ; thus  $g(x_1, x_2, x_r)$  is independent of  $x_r$  and we obtain

$$\int_{\mathcal{S}^{N-1}} cx_1^2 x_2^2 C_k d^E f_x d\mathcal{H}^{N-1}(x) = \int_{B_1(0)} cx_1^2 x_2^2 \int_{\sqrt{1-x_1^2-x_2^2}\mathcal{S}^{N-3}} d\mathcal{H}^{N-3}(x_r) d(x_1, x_2).$$

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  $x$ .



**Figure 10.** Nonlocal Hessian-based denoising with  $L^2$  data fidelity ( $p = 2$ ) on the real-world “flowers” image (top left) with  $\sigma = 0.1$  Gaussian noise. For this input, iterating the computation of the weights leads to a slight overregularization, clearly illustrating the piecewise quadratic structure of the model (bottom right).

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

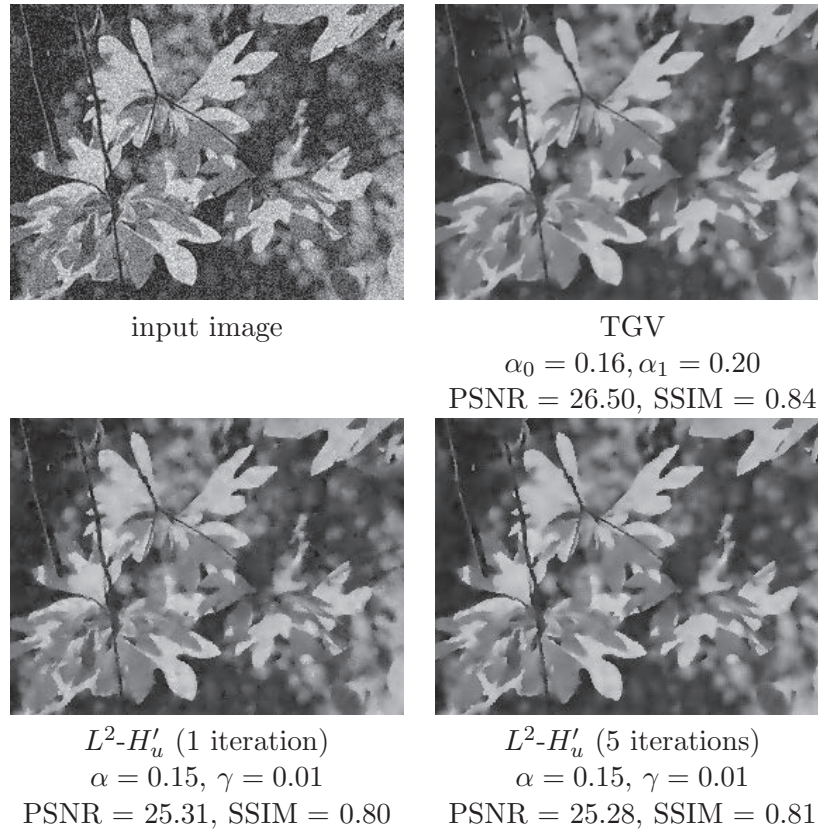
$$I_2 = EI_n E^\top = E \begin{pmatrix} B & x \end{pmatrix} \begin{pmatrix} B^\top \\ x^\top \end{pmatrix} E^\top = EBB^\top E + Exx^\top E^\top,$$

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

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



**Figure 11.** Nonlocal Hessian-based denoising with  $L^2$  data fidelity ( $p = 2$ ) on the real-world “leaves” image (top left) with  $\sigma = 0.1$  Gaussian noise (see also Figures 9 and 10).

Using the fact that  $c = C_k^{-1}$ , we obtain

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

The right-hand side is

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

$$\begin{aligned}
&= \frac{\pi}{4} |\mathcal{S}^{N-3}| \frac{8}{N(N+2)(N-2)} \\
&= \frac{\pi}{4} \frac{(N-2)\pi^{(N-2)/2}}{\Gamma(1+(N-2)/2)} \frac{8}{N(N+2)(N-2)} \\
&= \frac{2\pi^{N/2}}{N(N+2)\Gamma(N/2)}.
\end{aligned}$$

Using a suitable reordering of the coordinates, this proves the  $i_0 \neq j_0$  case in the claimed equality

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

The case  $i_0 = j_0$  follows by the same argument with  $g(x) = cx_1^4$ . The only difference lies in the evaluation of the integral in (5.3), which turns out to be a constant multiple of the first case, since

$$\int_0^{2\pi} \cos^4(\alpha) d\alpha = 3 \int_0^{2\pi} \cos^2 \alpha \sin^2 \alpha d\alpha.$$

This completes the proof. ■

## REFERENCES

- [AFP00] L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, New York, 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 J. Numer. Anal.*, 47 (2009), pp. 844–860.
- [BBM01] J. BOURGAIN, H. BREZIS, AND P. MIRONESCU, *Another look at Sobolev spaces*, in *Optimal Control and Partial Differential Equations: A Volume in Honour of A. Bensoussan's 60th Birthday*, IOS Press, Amsterdam, 2001, pp. 439–455.
- [BCM05] A. BUADES, B. COLL, AND J. M. MOREL, *A review of image denoising algorithms, with a new one*, *Multiscale Model. Simul.*, 4 (2005), pp. 490–530.
- [BEG08] A. BERTOZZI, S. ESEDOĞLU, AND A. GILLETTE, *Analysis of a two-scale Cahn–Hilliard model for binary image inpainting*, *Multiscale Model. Simul.*, 6 (2007), pp. 913–936.
- [Ber14] M. BERGOUNIOUX, *Mathematical analysis of a inf-convolution model for image processing*, *J. Optim. Theory Appl.*, DOI [10.1007/s10957-015-0734-8](https://doi.org/10.1007/s10957-015-0734-8), 2015.
- [BG04] A. L. BERTOZZI AND J. B. GREER, *Low-curvature image simplifiers: Global regularity of smooth solutions and Laplacian limiting schemes*, *Comm. Pure Appl. Math.*, 57 (2004), pp. 764–790.
- [BHS09] M. BURGER, L. HE, AND C.-B. SCHÖNLIEB, *Cahn–Hilliard inpainting and a generalization for grayvalue images*, *SIAM J. Imaging Sci.*, 2 (2009), pp. 1129–1167.
- [BKP10] K. BREDIES, K. KUNISCH, AND T. POCK, *Total generalized variation*, *SIAM J. Imaging Sci.*, 3 (2010), pp. 492–526.
- [BP10] M. BERGOUNIOUX AND L. PIFFET, *A second-order model for image denoising*, *Set-Valued Var. Anal.*, 18 (2010), pp. 277–306.
- [Bra02] A. BRAIDES,  *$\Gamma$ -Convergence for Beginners*, Oxford Lecture Ser. Math. Appl. 22, Oxford University Press, Oxford, UK, 2002.
- [Bre83] H. BREZIS, *Analyse fonctionnelle*, Vol. 5, Masson, Paris, 1983.
- [BV11] K. BREDIES AND T. VALKONEN, *Inverse problems with second-order total generalized variation constraints*, in *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*, J. Vis. Commun. Image R., 18 (2007), pp. 464–486.
- [CKS02] T. F. CHAN, S. H. KANG, AND J. SHEN, *Euler’s elastica and curvature-based inpainting*, SIAM J. Appl. Math., 63 (2002), pp. 564–592.
- [CL97] A. CHAMBOLLE AND P.-L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.
- [CMM01] T. CHAN, A. MARQUINA, AND P. MULET, *High-order total variation-based image restoration*, SIAM J. Sci. Comput., 22 (2000), pp. 503–516.
- [CS01] T. F. CHAN AND J. SHEN, *Nontexture inpainting by curvature-driven diffusions*, J. Vis. Commun. Image R., 12 (2001), pp. 436–449.
- [Dem85] F. DEMENGEL, *Fonctions à Hessien borné*, Ann. Inst. Fourier (Grenoble), 34 (1984), pp. 155–190.
- [DM93] G. DAL MASO, *Introduction to  $\Gamma$ -Convergence*, Birkhäuser, Boston, 1993.
- [DWB09] S. DIDAS, J. WEICKERT, AND B. BURGETH, *Properties of higher order nonlinear diffusion filtering*, J. Math. Imaging Vision, 35 (2009), pp. 208–226.
- [GO08] G. GILBOA AND S. OSHER, *Nonlocal operators with applications to image processing*, Multiscale Model. Simul., 7 (2008), pp. 1005–1028.
- [HS06] W. HINTERBERGER AND O. SCHERZER, *Variational methods on the space of functions of bounded Hessian for convexification and denoising*, Computing, 76 (2006), pp. 109–133.
- [KOJ05] S. KINDERMANN, S. OSHER, AND P. W. JONES, *Deblurring and denoising of images by nonlocal functionals*, Multiscale Model. Simul., 4 (2005), pp. 1091–1115.
- [LBL13] F. LENZEN, F. BECKER, AND J. LELLMANN, *Adaptive second-order total variation: An approach aware of slope discontinuities*, in Scale Space and Variational Methods in Computer Vision, Lecture Notes in Comput. Sci. 7893, Springer, Berlin, Heidelberg, 2013, pp. 61–73.
- [LBU12] S. LEFKIMMIATIS, A. BOURQUARD, AND M. UNSER, *Hessian-based norm regularization for image restoration with biomedical applications*, IEEE Trans. Image Process., 21 (2012), pp. 983–995.
- [LDM07] R. LERALLUT, E. DECENCIÈRE, AND F. MEYER, *Image filtering using morphological amoebas*, Image Vis. Comput., 25 (2007), pp. 395–404.
- [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 Trans. Image Process., 12 (2003), pp. 1579–1590.
- [LS11] G. LEONI AND D. SPECTOR, *Characterization of Sobolev and BV spaces*, J. Funct. Anal., 261 (2011), pp. 2926–2958.
- [LS14] G. LEONI AND D. SPECTOR, *Corrigendum to “Characterization of Sobolev and BV spaces,”* [J. Funct. Anal., 261 (2011), pp. 2926–2958], J. Funct. Anal., 266 (2014), pp. 1106–1114.
- [LT06] M. LYSAKER AND X. C. TAI, *Iterative image restoration combining total variation minimization and a second-order functional*, Int. J. Comput. Vis., 66 (2006), pp. 5–18.
- [LTC13] R. LAI, X.-C. TAI, AND T. F. CHAN, *A ridge and corner preserving model for surface restoration*, SIAM J. Sci. Comput., 35 (2013), pp. A675–A695.
- [Men12] T. MENGESHA, *Nonlocal Korn-type characterization of Sobolev vector fields*, Commun. Contemp. Math., 14 (2012), 1250028.
- [MM98] S. MASNOU AND J. M. MOREL, *Level lines based disocclusion*, in Proceedings of the IEEE International Conference on Image Processing, 1998, pp. 259–263.
- [MS15] T. MENGESHA AND D. SPECTOR, *Localization of nonlocal gradients in various topologies*, Calc. Var. Partial Differential Equations, 52 (2015), pp. 253–279.
- [Nir59] L. NIRENBERG, *On elliptic partial differential equations*, Ann. Scuola Norm. Sup. Pisa (3), 13 (1959), pp. 115–162.
- [OF03] S. OSHER AND R. FEDKIW, *Level Set Methods and Dynamic Implicit Surfaces*, Springer, New York, 2003.
- [PB13] K. PAPAITSOROS AND K. BREDIES, *A study of the one dimensional total generalised variation regularisation problem*, Inverse Probl. Imaging, 9 (2015), pp. 511–550.
- [Pon04a] A. C. PONCE, *An estimate in the spirit of Poincaré’s inequality*, J. Eur. Math. Soc. (JEMS), 6 (2004), pp. 1–15.
- [Pon04b] A. C. PONCE, *A new approach to Sobolev spaces and connections to  $\Gamma$ -convergence*, Calc. Var. Partial Differential Equations, 19 (2004), pp. 229–255.

- [PS08] C. PÖSCHL AND O. SCHERZER, *Characterization of minimizers of convex regularization functionals*, in *Frames and Operator Theory in Analysis and Signal Processing*, Contemp. Math. 451, AMS, Providence, RI, 2008, pp. 219–248.
- [PS13] C. PÖSCHL AND O. SCHERZER, *Exact solutions of one-dimensional total generalized variation*, *Commun. Math. Sci.*, 13 (2015), pp. 171–202.
- [PS14] K. PAPAITSOROS AND C. B. SCHÖNLIEB, *A combined first and second order variational approach for image reconstruction*, *J. Math. Imaging Vision*, 48 (2014), pp. 308–338.
- [PSS13] K. PAPAITSOROS, C. B. SCHÖNLIEB, AND B. SENGUL, *Combined first and second order total variation inpainting using split Bregman*, *Image Processing On Line*, 2013 (2013), pp. 112–136.
- [RBP14] R. RANFTL, K. BREDIES, AND T. POCK, *Non-local total generalized variation for optical flow estimation*, in *Proceedings of the International Conference on Computer Vision and Applications*, Springer, Berlin, 2014, pp. 439–454.
- [Rin00] W. RING, *Structural properties of solutions to total variation regularization problems*, *ESAIM Math. Model. Numer. Anal.*, 34 (2000), pp. 799–810.
- [ROF92] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, *Phys. D*, 60 (1992), pp. 259–268.
- [Sch98] O. SCHERZER, *Denoising with higher order derivatives of bounded variation and an application to parameter estimation*, *Computing*, 60 (1998), pp. 1–27.
- [Set99] J. A. SETHIAN, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, Cambridge Monogr. Appl. Comput. Math. 3, Cambridge University Press, Cambridge, UK, 1999.
- [SS08] S. SETZER AND G. STEIDL, *Variational methods with higher order derivatives in image processing*, in *Approximation Theory XII: San Antonio 2007*, Nashboro Press, Brentwood, TN, 2008, pp. 360–385.
- [SS14] T. SHIEH AND D. SPECTOR, *On a new class of fractional partial differential equations*, *Adv. Calc. Var.*, to appear; available online from <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*, *Commun. Math. Sci.*, 9 (2011), pp. 797–872.
- [WBSS04] Z. WANG, A. C. BOVIK, H. R. SHEIKH, AND E. P. SIMONCELLI, *Image quality assessment: From error visibility to structural similarity*, *IEEE Trans. Image Process.*, 13 (2004), pp. 600–612.
- [WBV11] M. WELK, M. BREUSS, AND O. VOGEL, *Morphological amoebas are self-snakes*, *J. Math. Imaging Vision*, 39 (2011), pp. 87–99.
- [Wel12] M. WELK, *Amoeba active contours*, in *Scale Space and Variational Methods in Computer Vision*, Lecture Notes in Comput. Sci. 6667, Springer, Berlin, 2012, pp. 374–385.