RNLp : Mixing Non-Local and TV-Lp methods to remove impulse noise from images

Julie Delon, Agnès Desolneux, Camille Sutour, Agathe Viano

Noname manuscript No.
(will be inserted by the editor)

RNLp : Mixing Non-Local and TV-Lp methods to remove impulse noise from images

Julie Delon · Agnès Desolneux · Camille Sutour · Agathe Viano

Received: date / Accepted: date

Abstract We propose a new variational framework to remove random-valued impulse noise from images. This framework combines, in the same energy, a nonlocal L^p data term and a total variation regularization term. The nonlocal L^p term is a weighted L^p distance between pixels, where the weights depend on a robust distance between patches centered at the pixels. In a first part, we study the theoretical properties of the proposed energy, and we show how it is related to classical denoising models for extreme choices of the parameters. In a second part, after having explained how to numerically find a minimizer of the energy thanks to primal-dual approaches, we show extensive denoising experiments on various images and noise intensities. The denoising performance of the proposed methods is on par with state of the art approaches, and the remarkable fact is that, unlike other successful variational approaches for impulse noise removal, they do not rely on a noise detector.

Keywords Image denoising · impulse noise · variational methods · patch-based methods · convex optimization

J. Delon, C. Sutour
MAP5, Université Paris Descartes.
E-mail: {julie.delon,camille.sutour}@parisdescartes.fr

A. Desolneux
CMLA, CNRS and ENS Paris-Saclay.
E-mail: agnes.desolneux@cmla.ens-cachan.fr

A. Viano
Deloitte France.
E-mail: viano.a@gmail.com

1 Introduction

Image denoising is one of the most studied inverse problems in image processing. Given a noisy version v of an original image f, image denoising aims at recovering f from the degraded observation v. In full generality, this inverse problem can only be managed thanks to well chosen priors on the noise distribution or on the regularity of the original image.

The different types of noise encountered in image processing result from random phenomena happening during acquisition and transmission steps of a digital image. In this paper, we focus on the case of *random-valued impulse noise*. Assuming that the original unobserved image $f : \Omega \to \mathbb{R}$ is defined on a discrete rectangular domain Ω, the degraded image v can be written

$$\forall i \in \Omega, \quad v(i) = (1 - b(i))f(i) + b(i)w(i),$$

where all the $b(i)$ and $w(i)$ are realizations of independent variables, following respectively a Bernoulli distribution of parameter $\rho \in [0, 1]$ (called the intensity of the impulse noise) for the $b(i)$, and a uniform distribution on the range of all possible gray-levels $\Gamma = [\gamma_0, \gamma_1]$ for the $w(i)$. Random-valued impulse noise can be due to errors in the transmission process (for instance with satellite images that are transmitted to the ground) or to errors in the acquisition process (when some captors are deficient or "jump" as it sometimes the case in resistivity borehole images, producing therefore random measurements).

Several strategies have been developed over the years to remove random-valued impulse noise. The simplest methods rely on the median and its extensions [51, 37], and modify all pixels indifferently. More evolved

methods usually make use of a well chosen noise detector to avoid oversmoothing non corrupted pixels, and combine it with a restoration method restricted to the set of corrupted pixels. The restoration approach can be based on filtering [25] or variational principles [47, 21], can be patch-based [56,31,18] or combine the last two approaches via dictionary learning [55,60], to name just a few. Several recent approaches are also dedicated to the mixture of Gaussian and impulse noise [33,56, 31,55,18]. In this paper, we propose to study a variational approach combining the total variation with a well designed non local "generalized" median.

Variational approaches. Variational methods for denoising an image v usually consist in the following optimization problem:

$$\underset{u\in\mathbb{R}^\Omega}{\operatorname{argmin}}\ E_{\text{data}}(u,v)+\lambda E_{\text{reg}}(u),$$

where E_{data} is related to the noise distribution, E_{reg} comes from a prior assumption on the solution u (and is thus related to the image formation model) and λ is a regularization weight balancing both terms. In the case of impulse noise reduction, a natural choice for E_{data} should rely on a L^0 term. In order to keep a convex formulation, it is usually advocated to rely on a L^1 fidelity term instead. One of the first attempts to remove impulse noise with a L^1 fidelity term is due to Nikolova [47] in 2004. She proposes to solve the following minimization problem

$$\underset{u\in\mathbb{R}^\Omega}{\operatorname{argmin}}\ \|u-v\|_1+\lambda\sum_i\phi(\nabla u(i)),$$

where ϕ is a convex and edge-preserving function, and where $\|u-v\|_1=\sum_i|u(i)-v(i)|$. The specific case of a L^1 norm for ϕ yields the famous (anisotropic) TV-L1 model

$$u_{\text{TV-L1}}\in\underset{u\in\mathbb{R}^\Omega}{\operatorname{argmin}}\ \|u-v\|_1+\lambda\,\text{TV}(u),$$

with $\text{TV}(u)=\sum_i\|\nabla u(i)\|_1$. This model, introduced in signal processing by Alliney [1] in 1992 and explored by Nikolova in image processing [46], has been studied extensively in the last fifteen years, both from the theoretical and numerical points of view [22]. Let us also mention that the same model can also be studied with a more isotropic total variation term, replacing $\|\nabla u(i)\|_1$ by $\|\nabla u(i)\|_2$ in the previous sum: See for instance the paper of Condat [15] in which the definition of the total variation for discrete images is discussed.

If the TV-L1 model performs well on edges and is simple to use in practice, it does not recover well textured areas, and it tends to produce piecewise constant regions in flat areas, a phenomenon which is known as the famous *staircasing effect*. Different directions have been proposed to improve this model for impulse noise removal. A first possibility is to reduce the number of pixels to be restored by relying on noise detectors. In 2004, Chan et al. [9] observe that the results can be widely improved by first detecting corrupted pixels and restricting the regularization to the detected locations. Several impulse noise detectors have been developed in the literature for this detection task, such as the Adaptive Center-Weighted median Filter (ACWMF) [12], the Rank Ordered Absolute differences (ROAD) [25] or the Rank Ordered Logarithmic Differences (ROLD) [20]. As pointed out by Duval in his PhD thesis [22], the Rank Ordered Absolute Differences (ROAD) usually yields the best detection performance in terms of stability and sensitivity to outliers.

Another direction of improvement, proposed by Duval in [22], consists in spatially adapting the value of λ. Duval proposes to use large values of λ for pixels detected as noisy, and very small values for the other pixels. The Total Variation hence plays its regularization role only on pixels that are suspected to be corrupted.

As mentioned before, the TV-L1 model can actually be seen as a convex relaxation of the nonconvex TV-L0 model [59]

$$u_{\text{TV-L0}}\in\underset{u\in\mathbb{R}^\Omega}{\operatorname{argmin}}\ \|u-v\|_0+\lambda\,\text{TV}(u),$$

where $\|u-v\|_0=\sum_i\mathbf{1}_{u(i)\neq v(i)}$ counts the nonzero entries of $u-v$. This model ensures sparsity but the minimization of this nonconvex, nonsmooth problem is known to be NP-hard [45]. Various solutions have been investigated to compute good approximate or exact solutions: convex relaxation [53,11] (hence the L^1 minimization problem), greedy algorithms such as Matching Pursuit and its variants [43], or continuous but nonconvex approximations such as the *Log-Sum penalty* [5], the *Smoothly Clipped Absolute Deviation* [23], or the L^p norms, $0<p<1$ [24]. The latter problems are still not convex, but a global minimizer can be computed using graph-cut techniques [38,34] or functional lifting [49, 50]. In the context of impulse noise removal, the TV-L0 model is studied by the authors of [59], who find a local minimum of the problem with a proximal ADMM. In a similar direction, the author of [57] proposes a formulation which finds simultaneously the damaged pixels and the restored image, with a TV data term on the image u, a sparsity L0 term on the set of noisy pixels and a quadratic data term on undamaged pixels.

Patch-based approaches. TV regularization works well to reconstruct an image consisting of smooth regions

and sharp edges, but it fails on fine textures and image details. In contrast, nonlocal approaches are very effective on repetitive patterns. These methods, introduced about ten years ago, led to very significant improvements in solving ill-posed inverse problems by combining repeated structures in images. The first denoising algorithms based on this idea appeared in 2005 with the NL-means [4] on the one hand, and UINTA filtering [2] and the DUDE algorithm [44] on the other hand. These nonlocal approaches have inspired in the last ten years a considerable number of works in image denoising, most of them being focused on Gaussian additive noise [35,36,39,61,58,54]. An enthralling and enlightening review of the recent advances on the subject is proposed in [40]. Different extensions to other noise models have been proposed, for instance in [16, 17].

The specific adaptation to impulse noise is the subject of several recent works. The most direct adaptation is a similar formulation to the NL-means developed in [10], leading to a Non-Local-Median formulation, solution of

$$u_{\text{NL-Med}} \in \operatorname*{argmin}_{u\in\mathbb{R}^\Omega} \sum_{i\in\Omega}\sum_{j\in\Omega} w_{i,j}|u(i)-v(j)|,$$

where the weights $w_{i,j}$ reflect the similarity between two patches centered at $v(i)$ and $v(j)$.

Several other papers adapt the NL-means approach to impulse noise removal. For instance in [41] and [31], the authors modify the trilateral filter of [25] and obtain a patch-based weighted means filter where the weights depend on the ROAD detector, the distance between the pixels and the similarity between patches (that is a L^2 distance weighted by the detector ROAD). In [56], the authors introduce a new "measure of outlyingness" and cluster pixels in groups depending on this measure. They rely on this measure to develop a coarse to fine strategy for impulse noise detection and then apply the NL-means using a "reference image". In [18], Delon and Desolneux propose a nonlocal approach that uses the maximum likelihood estimator instead of the mean or the median, and which measures the similarity between neighboring pixels thanks to a robust distance between patches. This method, called PARIGI, has good denoising performance (to test it online, see [19]) and does not use any noise detector.

Hybrid methods. The idea to combine nonlocal approaches and variational methods has been the subject of several works, often referred to as *hybrid methods*. An interesting review of such methods can be found in Section 6 of [52].

Inspired by the success of the NL-means algorithm for Gaussian noise removal, several variational interpretations of nonlocal methods have been proposed in the past ten years. In [26], Gilboa and Osher consider the nonlocal smoothness term $\sum_{i\in\Omega}\sum_{j\in\Omega}(u(i)-u(j))^2\omega_{i,j}$, and extend it to define the nonlocal total variation (NLTV) in [27]. It is then used to define a nonlocal version of the classical ROF (Rudin-Osher-Fatemi) model. In [32], Huang et. al use a nonlocal low rank regularizer in the regularization term to exploit the nonlocal self-similarity of natural images. In an orthogonal direction, Sutour et al. [52] propose a model where the nonlocality is introduced in the data term, while the regularization term is the total variation. This model is given by the energy

$$E_{\text{RNL2}}(u) = \sum_{i\in\Omega}\sum_{j\in\Omega} w_{i,j}(u(i)-v(j))^2 + \lambda\,\text{TV}(u),$$

where the weights are the nonlocal weights of NL-means. This is therefore equivalent to an adaptive TV regularization of the NL-means functional to restore images corrupted by Gaussian noise. The same authors propose a more general framework where the weighted L^2 term is replaced by the negative log-likelihood of the true pixel value given the observed noisy value, and their framework then applies to various noise models belonging to the exponential family: Gaussian, Gamma, Poisson, etc. Nonetheless, the impulse noise model requires a specific framework.

Contributions. In this paper, we explore strategies inspired by the recent work [52] to remove random-valued impulse noise from images. While some of the methods presented above [27,32] use nonlocal regularizers, we propose here to keep a classic TV regularization term and to express the nonlocality in the data fidelity term of the energy. Our model can be written as minimizing the energy

$$E_{\text{RNLp}}(u) = \sum_{i\in\Omega}\sum_{j\in\Omega} w_{i,j}|u(i)-v(j)|^p + \lambda\,\text{TV}(u), \quad (1)$$

with $p>0$, and where the $w_{i,j}$ are patch-based weights that measure the similarity between the patches centered respectively at i and j. We discuss several possible choices for the weights and for the power p, and we study the theoretical properties of the proposed energy. We also show which algorithms can be developed to find a minimizer of E_{RNLp}. Finally, we end with extensive experiments to show the performance and the limits of this proposed hybrid model. Let us emphasize that, unlike state-of-the-art impulse noise removal methods, this approach does not rely on any impulse noise detector.

2 A common variational framework

In this section, we discuss some of the theoretical properties of the energy E_{RNLp} and its possible links with similar energies.

2.1 The model: The Regularized Non-Local-L^p-Regression

Let $v : \Omega \to \mathbb{R}$ denote the noisy image. We propose here to study the hybrid model given by the energy (1).

where $\lambda \geq 0$ and where the weights $w_{i,j}$ are positive, but not necessarily normalized. These weights measure a "similarity" between the pixels i and j. The different possible choices for these weights will be discussed in details in the following section. For this section, we assume that the weights are fixed.

When $p \leq 1$, the energy might admit several minima. In the following, we will denote by u_{RNLp} one of the minimizers

$$u_{\mathrm{RNLp}} \in \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} E_{\mathrm{RNLp}}(u).$$

2.1.1 Case $\lambda = 0$ (NLp-Regression)

Let us first notice that if we take $\lambda = 0$, then the minimizers u_{NLp} of E_{RNLp} satisfy

$$u_{\mathrm{NLp}} \in \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j} |u(i) - v(j)|^p.$$

The complete energy E_{RNLp} is hence called the "Regularized Non-Local-L^p-Regression" energy, the regularization being performed thanks to the additional Total Variation term $\lambda\,\mathrm{TV}(u)$.

When $p = 1$, the Non-Local-L^p-Regression solutions are given by usual median values:

$$\begin{aligned} \forall i \in \Omega, \quad u_{\mathrm{NL1}}(i) &= u_{\text{NL-Med}}(i) \\ &= \mathrm{Median}\{(v(j), w_{i,j}), j \in \Omega\}. \end{aligned}$$

Here, assuming the x_j are increasingly ordered, the notation $\mathrm{Median}\{(x_j, p_j), 1 \leq j \leq J\}$ corresponds to any value y in the interval $[x_{j_0}, x_{j_0+1}]$ where j_0 is such that $\sum_{j=1}^{j_0} p_j \leq \frac{1}{2} \sum_{j=1}^{J} p_j \leq \sum_{j=1}^{j_0+1} p_j$. The solution u_{NL1} turns out to be exactly the Non-Local-Median introduced in [10].

When $p = 2$, the NL2-Regression is simply the weighted mean, and therefore the solutions are given by

$$\begin{aligned} \forall i \in \Omega, \quad u_{\mathrm{NL2}}(i) &= \mathrm{Mean}\{(v(j), w_{i,j}), j \in \Omega\} \\ &= \frac{\sum_j w_{i,j} v(j)}{\sum_j w_{i,j}} = u_{\text{NL-Means}}(i), \end{aligned}$$

and turns out to be exactly the solution computed by the NL-means algorithm, for the appropriate choice of the weights $w_{i,j}$ [4].

When $p = 0$, the solutions are given by the modes of the set of values $v(j)$, weighted by the values $w_{i,j}$,

$$\forall i \in \Omega, \quad u_{\mathrm{NL0}}(i) = \mathrm{Mode}\{(v(j), w_{i,j}), j \in \Omega\}.$$

Here, the notation $\mathrm{Mode}\{(x_j, p_j), 1 \leq j \leq J\}$ means the value x_{j_0} such that p_{j_0} is the max of the p_j. Such a j_0 is not necessarily unique. This mode estimator is closely related to the method developed in the patch-based approach PARIGI [18]. Indeed, the authors of PARIGI estimated a denoised value at pixel i as being the mode of the empirical histogram of the values $v(j)$ (only the n values with highest weights w_{ij} were kept), convolved with the probability density of the noise model (a mixture of uniform and Gaussian law). Here, in the NL0 approach, the framework is a bit different: the weights w_{ij} are taken into account, and there is no convolution with a parametric distribution.

2.1.2 RNLp or NLp-TV?

The authors of [52] consider the minimization problem

$$\begin{aligned} u_{\mathrm{RNL2}} &= \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} \; E_{\mathrm{RNL2}}(u) \\ &= \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j} (u(i) - v(j))^2 \; + \lambda\,\mathrm{TV}(u). \end{aligned}$$

An elementary computation shows that this solution can also be written as

$$\begin{aligned} u_{\mathrm{RNL2}} &= \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} \; E_{\text{NL2-TV}}(u) \\ &= \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} \sum_{i \in \Omega} w_i \left(u(i) - u_{\mathrm{NL2}}(i)\right)^2 \; + \lambda\,\mathrm{TV}(u), \end{aligned}$$

where $w_i = \sum_j w_{i,j}$. The two energies E_{RNL2} and $E_{\text{NL2-TV}}$ are equal up to a constant independent of u (it just depends on v and on the weights w). Therefore, as noticed by Sutour *et al.* in [52], finding a minimizer of E_{RNL2} is equivalent to applying a spatially varying (with adaptive weights) TV-L2 regularization to the Non-Local means u_{NL2}. We can wonder if the situation is similar in our framework. Indeed, in a similar way, replacing the L^2 norm by the L^p one, we can consider the TV-Lp regularization of the Non-Local-L^p-Regression u_{NLp}, that is the energy

$$E_{\text{NLp-TV}}(u) = \sum_{i \in \Omega} w_i |u(i) - u_{\mathrm{NLp}}(i)|^p + \lambda\,\mathrm{TV}(u). \quad (2)$$

Now, when $p \neq 2$, this energy can generally not be written as E_{RNLp} plus a constant, and more generally one can show that the two energies may not admit the

same minimizers, even for different values of λ. This is the aim of the following proposition that presents an example. Let us however remark that, when $\lambda = 0$ the two energies have the same minimizers $u = u_{\text{NLp}}$.

Proposition 1 *When $p \neq 2$, the two energies E_{RNLp} (1) and $E_{\text{NLp-TV}}$ (2) do not necessarily have the same minimizers.*

To prove this proposition, a numerical counter-example suffices in that case. It is postponed to the Appendix.

The above theoretical difference between the minimizers of the two energies can also be checked from an experimental viewpoint. This is illustrated on Figure 1, with $p = 1$. For the same value of λ, the minimizers of the energies E_{RNL1} and $E_{\text{NL1-TV}}$ are clearly different. We have also tried using different values of λ for the second energy $E_{\text{NL1-TV}}$. The minimizer, for an optimal choice of λ becomes very similar to the one of the first energy, but is however still different.

2.2 The choice of the weights

In this section we discuss some possible choices for the weights in E_{RNLp}.

A first very simple choice would be

$$w^{\delta}_{i,j} = \mathbb{1}_{i=j}.$$

In this case, the energy E_{RNLp} boils down to a simple TV-L^p model.

A second possibility considered in this paper is to use patch-based weights. To this aim, we need to define a distance between patches which remains robust in the presence of impulse noise. In the following, we write P_i for a square patch in v centered at pixel i. In the work [18], the different possible distances between patches were studied in details. The conclusion was that a distance able to take small values only for similar patches despite the presence of impulse noise was given by

$$d(P_i, P_j)^2 = \sum_{k=1}^{(2s+1)^2} \mathcal{B}((2s+1)^2, k, (1-\rho)^2)\ |P_i - P_j|^2_{(k)}, \tag{3}$$

where $(2s+1)^2$ is the number of pixels in the patches (the half-patch size is s), $|P_i - P_j|_{(1)} \leq |P_i - P_j|_{(2)} \leq \cdots \leq |P_i - P_j|_{(2s+1)^2}$ are the values obtained by ordering the $(2s+1)^2$ values of the differences $|P_i(z) - P_j(z)|$ for $z \in [-s, s]^2$, ρ is the impulse noise intensity and $\mathcal{B}$ denotes the tail of the binomial distribution given for all $0 \leq k \leq n$ integers and $q \in [0, 1]$ by

$$\mathcal{B}(n, k, q) = \sum_{i=k}^{n} \binom{n}{i} q^i (1-q)^{n-i}.$$

Notice in particular that when $\rho = 0$, the distance $d(P_i, P_j)$ is simply the L^2 distance between the two patches. The idea underlying the distance in Equation (3) is the following property: if P and Q are two independent random patches obtained by adding impulse noise to the same original patch $P^0 = Q^0$, the probability that the k^{th} difference $|P - Q|_{(k)}$ stems from two untouched pixels is $\mathcal{B}((2s+1)^2, k, (1-\rho)^2)$ (with the approximation that the smallest distances correspond to untouched pixels). Observe that this distance requires to know or to estimate the impulse noise intensity ρ. In this paper, for the sake of simplicity, we assume that ρ is known, but a rather precise estimation could be obtained with a noise estimator [18]. From the robust distance between patches, one can obtain different possible weights. We will here consider three of them:

- In a way similar to the NL-means, replacing the L^2 distance by the robust distance, we can define the weights

$$\widetilde{w}^{P}_{i,j} = \frac{1}{Z_i} e^{-d(P_i,P_j)^2/2h^2}, \tag{4}$$

  where $h > 0$ is a filtering parameter that has to be tuned, and $Z_i = \sum_j e^{-d(P_i,P_j)^2/2h^2}$ is a normalizing factor that ensures that $\sum_j \widetilde{w}^{P}_{i,j} = 1$.
- Now, there is no obligation to have normalized weights. Indeed, some patches may be rare, in the sense that there are not many patches that are close to them for the distance d, and we may not want to take them too much into account. Therefore, we can also consider the nonnormalized weights given by

$$w^{P}_{i,j} = e^{-d(P_i,P_j)^2/2h^2}. \tag{5}$$

- Finally, as it is also sometimes done in variants of NL-means, we can consider the n nearest patches to P_i for the distance d, and denoting by $V^d_n(i)$ the set of central pixels of these patches, we can define

$$w^{NN}_{i,j} = \mathbb{1}_{j \in V^d_n(i)}.$$

Note that in each case, the computation of the weights can be restricted to a search window $\mathcal{N}_i$ centered around each pixel i, by setting $w_{i,j} = 0$ for $j \notin \mathcal{N}_i$. This allows to reduce the amount of computations, as well as the number of possibly not relevant candidates. In practice, we always restrict the search window to a square neighborhood of half-size 7, while the patch half-size is always set to $s = 3$. The different choices for the weights will be further discussed in Section 4.

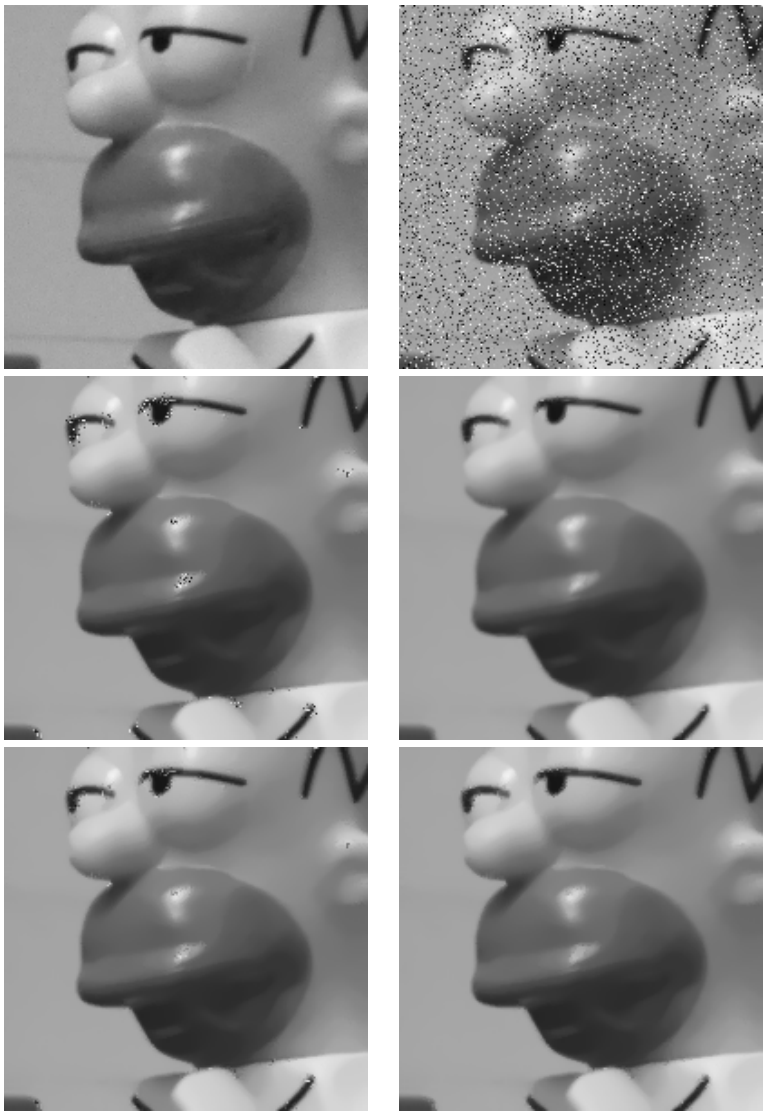

Fig. 1 Comparison between E_{RNL1} and $E_{\text{NL1-TV}}$. First line: left, the original image (a part of the *Simpson* image); right, the noisy image v, with an impulse noise of intensity $\rho = 0.2$. Second line: left, the NL-Median u_{NL1} obtained with weights w_{ij} of Equation (4) (see next subsection) with $h = 0.8$; right, minimizer u_{RNL1} of the energy E_{RNL1} with $\lambda = 0.5$. Third line: left, minimizer $u_{\text{NL1-TV}}$ of the energy $E_{\text{NL1-TV}}$ with $\lambda = 0.5$ and right, minimizer $u^{(0.95)}_{\text{NL1-TV}}$ of the energy $E_{\text{NL1-TV}}$ with $\lambda = 0.95$. This value of λ is the one such that $\|u^{(\lambda)}_{\text{NL1-TV}} - u_{\text{RNL1}}\|_2$ is the smallest. The PSNR values are 30.9 for u_{NL1}, 37.22 for u_{RNL1}, 34.63 for $u_{\text{NL1-TV}}$, and 36.87 for $u^{(0.95)}_{\text{NL1-TV}}$.

3 Algorithms

In this section, we present different algorithms that can be used to find a minimizer of the energy E_{RNLp} given by Equation (1) when $p \leq 1$. For $p > 1$, the energy is the sum of a differentiable function and the total variation, and can be minimized with a forward-backward splitting algorithm [13]. However, choosing $p > 1$ in the data term makes little sense in the presence of impulse noise. For $p = 1$, the formulation is nonsmooth but still convex. Such problems can be efficiently minimized thanks to augmented Lagrangian type techniques such as the Alternative directions method of multipliers (ADMM) [28,3] or the split-Bregman algorithm [29], or using the primal-dual algorithm of Chambolle and Pock [7], although some work is necessary here to compute the proximal operator of the nonlocal data term. For $p \in (0, 1)$ though, the formulation is not convex anymore and its minimization requires more tools. When $p = 1$, observe that the energy E_{RNLp} might have several global minima, and several local ones when $p < 1$.

3.1 Case $p = 1$

When $p = 1$, we rely on the primal-dual algorithm of Chambolle and Pock [7] to minimize E_{RNLp} efficiently. We first recall the general setting of this algorithm, then we explain how it can be adapted to our problem. The energy we deal with can be written as

$$E(u) = G(u) + F(Ku),$$

where in our setting: $u \in \mathbb{R}^{|\Omega|}$, $Ku = \lambda\nabla u$, $F(q) = \sum_i |q(i)|$ for $q \in \mathbb{R}^{|\Omega|} \times \mathbb{R}^{|\Omega|}$ (since we use the anisotropic version of the total variation), and $G(u) = \sum_{i\in\Omega}\sum_{j\in\mathcal{N}_i} w_{i,j}|u(i)-v(j)|$. For the discrete gradient, we use finite differences: for a discrete image u of size $n_r \times n_c$,

$$\nabla u^{(1)}(i_x, i_y) = \begin{cases} u(i_x+1, i_y) - u(i_x, i_y) & \text{if } i_x < n_r \\ 0 & \text{else} \end{cases}$$

$$\nabla u^{(2)}(i_x, i_y) = \begin{cases} u(i_x, i_y+1) - u(i_x, i_y) & \text{if } i_y < n_c \\ 0 & \text{else.} \end{cases}$$

In the following, we will denote by F^* the convex conjugate of F, given by

$$F^*(q) = \chi_{\mathcal{C}}(q) = \begin{cases} 0 & \text{if } q \in \mathcal{C}, \\ +\infty & \text{if } q \notin \mathcal{C}, \end{cases}$$

where

$$\mathcal{C} = \{q = (q^{(1)}, q^{(2)}) \in \mathbb{R}^{|\Omega|} \times \mathbb{R}^{|\Omega|} \ ; \\ |q^{(k)}(i)| \le 1 \ \forall k \in \{1,2\}, \ \forall i \in \Omega\}.$$

Let K^* denote the dual operator of the linear operator $K = \lambda\nabla$, given by

$$K^*(q) = -\lambda \text{div}(q),$$

where div is the (discrete) divergence operator, defined as the adjoint of the discrete gradient:

$$\text{div}(q)(i_x, i_y) = \\ \begin{cases} q^{(1)}(i_x, i_y) - q^{(1)}(i_x-1, i_y) & \text{if } 1 < i_x < n_r \\ q^{(1)}(i_x, i_y) & \text{if } i_x = 1 \\ -q^{(1)}(i_x-1, i_y) & \text{if } i_x = n_r \end{cases} \\ + \begin{cases} q^{(2)}(i_x, i_y) - q^{(2)}(i_x, i_y-1) & \text{if } 1 < i_y < n_c \\ q^{(2)}(i_x, i_y) & \text{if } i_y = 1 \\ -q^{(2)}(i_x, i_y-1) & \text{if } i_y = n_c \end{cases}.$$

Our convex but nonsmooth energy can be minimized thanks to the algorithm of Chambolle and Pock, summarized in Algorithm 1. This algorithm has been proved to converge [7], provided that the step sizes satisfy $\sigma\tau L^2 \le 1$, where $L = \|K\|$ is the norm of the gradient operator (see [14]). Given such step sizes, we can use the primal-dual residual as a stopping criterion, putting an end to the algorithm when the primal-dual residual gets below a certain threshold $\varepsilon = 10^{-4}$.

Algorithm 1 Chambolle-Pock Primal-Dual Algorithm

Parameters: $\tau, \sigma > 0$
Initialization: $\overline{u}^0 = u^0 = 0, q^0 = 0$
repeat
  Dual Step: $q^{n+1} = \text{prox}_{\sigma F^*}(q^n + \sigma K\overline{u}^n)$;
  Primal Step: $u^{n+1} = \text{prox}_{\tau G}(u^n - \tau K^* q^{n+1})$;
  $\overline{u}^{n+1} = 2u^{n+1} - u^n$;
  $n = n + 1$;
until convergence

This algorithm uses proximal operators. We recall that they are defined in the following way. For $\varphi : \mathbb{R}^N \to \mathbb{R}$, its proximal operator is $\text{prox}_\varphi : \mathbb{R}^N \to \mathbb{R}^N$ given by

$$\text{prox}_\varphi(x) = \underset{y\in\mathbb{R}^N}{\text{argmin}}\, \varphi(y) + \frac{1}{2}\|y - x\|_2^2.$$

In particular, since F^* is the indicator function of a convex set, $\text{prox}_{\sigma F^*}$ is the orthogonal projection on this set for any $\sigma > 0$:

$$\text{prox}_{\sigma F^*}(q)_i = \text{prox}_{F^*}(q)_i \\ = \left(\frac{q^{(1)}(i)}{\max(1, |q^{(1)}(i)|)}, \frac{q^{(2)}(i)}{\max(1, |q^{(2)}(i)|)}\right).$$

Now, we also need to compute the proximal operator of G that is a sum of L^1 terms. When there is just one term in the sum, the proximal operator is well-known, it is the *soft-thresholding*:

$$\forall \tau > 0, \forall x \in \mathbb{R}, \ \ \text{prox}_{\tau|\cdot|}(x) = \max\left(0, 1 - \frac{\tau}{|x|}\right) x,$$

which can also be reformulated as a median of three values

$$\forall \tau > 0, \forall x \in \mathbb{R}, \ \ \text{prox}_{\tau|\cdot|}(x) = \text{Median}\{0, x - \tau, x + \tau\}.$$

This result can be generalized for a sum of L^1 norms, as shown by the following proposition.

Proposition 2 *Assume that the data v and the weights w are given. Let $G : \mathbb{R}^{|\Omega|} \to \mathbb{R}$ be*

$$G(u) = \sum_{i\in\Omega}\sum_{j\in\mathcal{N}_i} w_{i,j}|u(i) - v(j)|,$$

where for all i, $\mathcal{N}_i$ is a set of pixels j such that $w_{i,j} > 0$. For each i, assuming that $\mathcal{N}_i$ contains J pixels, we sort the $\{v(j)\}_{j\in\mathcal{N}_i}$ in a sequence

$$v_{j_1} \le \cdots \le v_{j_J}.$$

Then, for all $\tau > 0$, the proximal operator of τG is:

$$\forall u \in \mathbb{R}^{|\Omega|}, \quad \mathrm{prox}_{\tau G}(u)_i = \\ \mathrm{Median}\{v_{j_1}, \dots, v_{j_J}, u(i) + \tau W_0, \dots, u(i) + \tau W_J\},$$

where Median *denotes the usual median value of a set (no weights), and where*

$$\forall k = 0, \dots, J, \quad W_k = -\sum_{l \le k} w_{i,l} + \sum_{k+1 \le l \le J} w_{i,l}.$$

Proof: The proof is an adaptation of a more general result shown by Li and Osher in [42]. Their result is more generic since it gives a formula to find the minimizer of a function that is a weighted sum of L^1 terms plus a differentiable strictly convex function. For the sake of completeness, we still provide the proof in our simpler framework.

Let us first notice that $W_k - W_{k-1} = -2w_{i,k} < 0$ and therefore

$$W_J < W_{J-1} < \dots < W_0.$$

Computing the i^{th} coordinate of the prox of τG, with $\tau \in \mathbb{R}_+$, at u is equivalent to find

$$y^* = \underset{y \in \mathbb{R}}{\mathrm{argmin}} \frac{1}{2}(u(i) - y)^2 + \tau \sum_{l=1}^{J} w_{i,l} |y - v_{j_l}|.$$

Let us denote $y \mapsto A_{u,\tau G}(y)$ the right-hand side function. It is a strictly convex function of y and there is thus a unique minimizer y^*.

We consider two cases. The first case is $y^* \notin \{v_{j_1}, \dots, v_{j_J}\}$. Then, $A_{u,\tau G}$ is differentiable at y^* and its derivative is equal to 0 at this point. This implies that

$$0 = y^* - u(i) + \tau \sum_{l=1}^{J} w_{i,l} \operatorname{sgn}(y - v_{j_l}).$$

Let us denote by $l^* \in \{0, 1, \dots, J\}$ the integer such that

$$v_{j_0} < v_{j_1} < \dots < v_{j_{l^*}} < y^* < v_{j_{l^*+1}} < \dots < v_{j_J}, \qquad (6)$$

with the convention that $v_{j_0} = -\infty$ and $v_{j_{J+1}} = +\infty$. This implies that

$$y^* = u(i) - \tau \sum_{l=1}^{J} w_{i,l} \operatorname{sgn}(y - v_{j_l}) = u(i) + \tau W_{l^*}.$$

And therefore

$$u(i) + \tau W_J < \dots < u(i) + \tau W_{l^*+1} < y^* = \\ u(i) + \tau W_{l^*} < u(i) + \tau W_{l^*-1} < \dots < u(i) + \tau W_0. \qquad (7)$$

Combining Equations (6) and (7), we conclude that y^* belongs to the set $\{v_{j_1}, \dots, v_{j_J}, u(i) + \tau W_0, \dots, u(i) + \tau W_J\}$, that is has $J+1$ values below it and $J+1$ values above it in this set. This shows that

$$y^* = \mathrm{Median}\{v_{j_1}, \dots, v_{j_J}, u(i) + \tau W_0, \dots, u(i) + \tau W_J\}.$$

Now, we have to consider a second case: there exists $l^* \in \{1, \dots, J\}$ such that $y^* = v_{j_{l^*}}$. Then $A_{u,\tau G}$ is not differentiable at y^*, but it admits a subdifferential at this point. The condition that y^* is the point of minimum of $A_{u,\tau G}$ becomes

$$0 \in \partial A_{u,\tau G}(y^*) \\ = y^* - u(i) + \tau \sum_{k \neq l^*} w_{i,k} \operatorname{sgn}(y - v_{j_k}) + \tau w_{i,l^*}[-1, 1].$$

This is equivalent to

$$u(i) + W_{l^*} < y^* = v_{j_{l^*}} < u(i) + W_{l^*-1}.$$

As in the first case we again conclude that y^* belongs to the set $\{v_{j_1}, \dots, v_{j_J}, u(i) + \tau W_0, \dots, u(i) + \tau W_J\}$, and that it is moreover the median value of this set. □

3.2 Case $p \in (0, 1)$

Minimizing E_{RNLp} for $p \in (0, 1)$ is not obvious since the energy is not convex anymore. Nonetheless, the energy belongs to the class of minimization problems with a nonconvex data term and TV regularization, and global solutions can be computed by lifting the problem to a higher dimensional space where the formulation becomes convex, following the ideas of [50,49,6]. We recall in the following the main steps of this approach in a spatially continuous setting.

3.2.1 A convex formulation of a nonconvex problem

Consider the continuous variational problem

$$\min_u G(u) + \lambda \, \mathrm{TV}(u) \\ := \quad \min_u \int_{\Omega_0} g(u(x), x) + \lambda |\nabla u(x)| \mathrm{d}x, \qquad (8)$$

where $u : \Omega_0 \to \Gamma_0$ is an image defined on a continuous image domain $\Omega_0 \subset \mathbb{R}^2$, $\Gamma_0 = [\gamma_0, \gamma_1]$ is the range of image intensities and $g : \Gamma_0 \times \Omega_0 \to \mathbb{R}^+$ is a non-negative, nonconvex function. It can be seen as the continuous analog to (1) where we note $g(u(x), x) = \int_{\Omega_0} w(x, y)|u(x) - v(y)|^p \mathrm{d}y$.

The key idea to deal with such a nonconvex formulation is to lift the functional using the level set representation given as follows:

$$\phi : \Omega_0 \times \Gamma_0 \to \{0, 1\}$$

$$\phi(x, \gamma) = \mathbf{1}_{u > \gamma}(x) = \begin{cases} 1, \text{ if } u(x) > \gamma, \\ 0, \quad \text{otherwise.} \end{cases}$$

Algorithm 2 Primal-Dual Algorithm for (11)

Parameters: $\sigma, \tau > 0$
Initialization: $\phi^0 = \bar{\phi}^0 = 0,\ q^0 = 0$
repeat
$\phi^{k+1} = \mathcal{P}_C\left(\phi^k + \tau \mathrm{div} q^k\right)$
$q^{k+1} = \mathcal{P}_K\left(q^k + \sigma \nabla \bar{\phi}^{k+1}\right)$
$\bar{\phi}^{k+1} = 2\phi^{k+1} - \phi^k$
until convergence.
return $\phi^* = \phi^{k+1}$, compute the binary solution $\mathbf{1}_{\phi^* > \mu}$ and the final estimate u^* using (10).

The authors of [50] showed that problem (8) is equivalent to the following problem:

$$\min_{\phi \in D} \int_{\Omega_0 \times \Gamma_0} \lambda |\nabla_x \phi(x,\gamma)| + g(\gamma, x) |\partial_\gamma \phi(x,\gamma)| \, \mathrm{d}x \mathrm{d}\gamma, \quad (9)$$

where $\nabla_x \phi$ denotes the spatial bi-dimensional gradient of ϕ, and $D = \{\phi : \Omega_0 \times \Gamma_0 \to \{0,1\} \mid \phi(x,\gamma_0) = 1 \text{ and } \phi(x,\gamma_1) = 0\}$ is the feasible set of such level set functions ϕ.

Note that the original image u can be recovered from ϕ using the following formula:

$$u(x) = \gamma_0 + \int_{\Gamma_0} \phi(x,\gamma) \, \mathrm{d}\gamma. \quad (10)$$

Interestingly here, the nonconvex function g in (9) is now only a function of (γ, x) and not $(\gamma, u(x))$, thus this formulation gets rid of the nonconvex part in u. This problem is however still not convex, due to the constraint $\phi \in D$. Hence, one can relax the problem and consider the following convex formulation:

$$\min_{\phi \in C} \int_{\Omega_0 \times \Gamma_0} \lambda |\nabla_x \phi(x,\gamma)| + g(\gamma, x) |\partial_\gamma \phi(x,\gamma)| \, \mathrm{d}x \, \mathrm{d}\gamma, \quad \text{with} \quad (11)$$

$$C = \{\phi : [\Omega_0 \times \Gamma_0] \to [0,1] \mid \phi(x,\gamma_0) = 1 \text{ and } \phi(x,\gamma_1) = 0\}.$$

This time the set C is convex, so the global minimizer of problem (11) can be computed. It is then shown in [50] that for a given $\phi^* \in C$ solution of (11), a global minimizer of the binary problem (9) can be obtained by computing the characteristic function $\mathbf{1}_{\phi^* > \mu}$ for almost every threshold $\mu \in [0,1]$.

3.2.2 Implementation

Problem (11) can be written equivalently in the following primal-dual formulation:

$$\min_{\phi \in C} \max_{q \in K} \int_{\Omega_0 \times \Gamma_0} \nabla \phi \cdot q,$$

where ∇ denotes the three-dimensional gradient and

$$K = \left\{ q = \left(q^{(1)}, q^{(2)}, q^{(3)}\right) : \Omega_0 \times \Gamma_0 \to \mathbb{R}^3 \mid |q^{(1)}| \le \lambda, |q^{(2)}| \le \lambda, \text{ and } |q^{(3)}| \le g \right\}.$$

Note that the inequalities are meant point-wise, e.g., for all $(x,\gamma) \in \Omega_0 \times \Gamma_0, |q_3(x,\gamma)| \le g(x,\gamma)$. Based on this formulation, one can then derive a primal-dual iteration scheme where the updates for the primal and dual variables are given in Algorithm 2. The projections $\mathcal{P}_C$ and $\mathcal{P}_K$ are respectively the projections onto the sets C and K. The former is simply obtained by truncating the iterate ϕ to the interval $[0,1]$ and by setting $\phi(.,\gamma_0) = 1$ and $\phi(.,\gamma_1) = 0$. The latter is a point-wise projection of the dual variable q computed as follows:

$$\mathcal{P}_K(q) = \left(\frac{q^{(1)}}{\max\left(\frac{|q^{(1)}|}{\lambda}, 1\right)}, \frac{q^{(2)}}{\max\left(\frac{|q^{(2)}|}{\lambda}, 1\right)}, \frac{q^{(3)}}{\max\left(\frac{|q^{(3)}|}{g}, 1\right)} \right).$$

Discretization. Similarly to the case $p = 1$, we use finite differences to define the discrete gradient and divergence operators. We define the lifted image ϕ on a discrete Cartesian grid of size $n_r \times n_c \times n_t$, and we get the following discrete gradient operator:

$$\nabla\phi^{(1)}(i_x,i_y,i_z) = \begin{cases} \frac{\phi(i_x+1,i_y,i_z) - \phi(i_x,i_y,i_z)}{h_r} & \text{if } i_x < n_r, \\ 0 & \text{else.} \end{cases}$$

$$\nabla\phi^{(2)}(i_x,i_y,i_z) = \begin{cases} \frac{\phi(i_x,i_y+1,i_z) - \phi(i_x,i_y,i_z)}{h_c} & \text{if } i_y < n_c, \\ 0 & \text{else.} \end{cases}$$

$$\nabla\phi^{(3)}(i_x,i_y,i_z) = \begin{cases} \frac{\phi(i_x,i_y,i_z+1) - \phi(i_x,i_y,i_z)}{h_t} & \text{if } i_z < n_t, \\ 0 & \text{else.} \end{cases}$$

Then the discrete divergence operator is computed as follows:

$$div(q)_{(i_x,i_y,i_z)} =$$
$$\begin{cases} \frac{q^{(1)}(i_x,i_y,i_z) - q^{(1)}(i_x-1,i_y,i_z)}{h_r} & \text{if } 1 < i_x < n_r \\ \frac{q^{(1)}(i_x,i_y,i_z)}{h_r} & \text{if } i_x = 1 \\ -\frac{q^{(1)}(i_x-1,i_y,i_z)}{h_r} & \text{if } i_x = n_r \end{cases}$$
$$+ \begin{cases} \frac{q^{(2)}(i_x,i_y,i_z) - q^{(2)}(i_x,i_y-1,i_z)}{h_c} & \text{if } 1 < i_y < n_c \\ \frac{q^{(2)}(i_x,i_y,i_z)}{h_c} & \text{if } i_y = 1 \\ -\frac{q^{(2)}(i_x,i_y-1,i_z)}{h_c} & \text{if } i_y = n_c. \end{cases}$$
$$+ \begin{cases} \frac{q^{(3)}(i_x,i_y,i_z) - q^{(3)}(i_x,i_y,i_z-1)}{h_t} & \text{if } 1 < i_z < n_t \\ \frac{q^{(3)}(i_x,i_y,i_z)}{h_t} & \text{if } i_z = 1 \\ -\frac{q^{(3)}(i_x,i_y,i_z-1)}{h_t} & \text{if } i_z = n_t. \end{cases}$$

For the spatial components of the gradient it is common to use discretization steps $h_r = h_c = 1$, however for the third coordinate it depends on the discretization of the intensity range. When dealing with a third dimension, for example time, it is not always clear how to deal with regularity and discretization in this third dimension [30]. In our case, we have chosen to consider $n_t = 256$ gray levels and a discretization step $h_t = \frac{1}{100}$ which matches the spatial regularity and allows for a regularization parameter λ of the same range as in the convex formulation when $p = 1$. Note that in order to accelerate the computation, one can subsample the intensity range by a factor κ, in which case the discretization step then becomes $h_t \times \kappa$.

Convergence and acceleration. The convergence of the algorithm has been proven provided that the stepsizes in Algorithm 2 fulfill the condition $\sigma\tau L^2 \leq 1$, where $L = \|\nabla\|$ is the norm of the (three-dimensional) gradient operator. The convergence is assumed when the primal-dual residual gets below a certain convergence tolerance $\varepsilon = 5.10^{-4}$.

However, since in this specific problem the data term is locally adaptive, based on the nonlocal weights, the value of the data-term g is greatly spatially-dependent. This makes the constraint on the dual variable q really local, which results in a nonspatially uniform convergence. Indeed, we have observed that the primal-dual residual might be really close to zero in most areas, while in some localized regions, where the data term g is high, a satisfactory convergence might not be reached.

To remedy this behavior, we have investigated the use of diagonal preconditioning [48,8]. We denote by n the number of elements of the variable ϕ, i.e. $n = n_c \times n_r \times n_t$, and $m = 3n$ the size of the dual variable. Then, given a sequence of stepsizes $(\tilde{\tau}_i)_{1\leq i\leq n}$ and $(\tilde{\sigma}_j)_{1\leq j\leq m}$ and $\alpha \in [0, 2]$, the diagonal matrices $T = \mathrm{diag}(\tau_i)$ and $\Sigma = \mathrm{diag}(\sigma_j)$, with

$$\tau_i = \frac{\tilde{\tau}_i}{\sum_{j=1}^{m} \tilde{\sigma}_j |\nabla_{j,i}|^{2-\alpha}}, \text{ and } \sigma_j = \frac{\tilde{\sigma}_j}{\sum_{i=1}^{n} \tilde{\tau}_i |\nabla_{j,i}|^{\alpha}},$$

satisfy the convergence criterion $\|\Sigma^{1/2}\nabla T^{1/2}\| \leq 1$. Note that the case $\nabla_{j,i} = 0$ is not an issue in practice, since it leads in Algorithm (2) to no gradient update for the dual variable q, so the step σ_j can be arbitrarily chosen in this case.

In order to take into account the local adaptivity of our problem, we use $\alpha = 1$ and we have found that the following sequences provide an efficient accelerated and uniform convergence:

$$\tilde{\tau} = 1_{\mathbb{R}^n}, \quad \text{and} \quad \tilde{\sigma} = [1_{\mathbb{R}^n}\ 1_{\mathbb{R}^n}\ \mathbf{g}],$$

where $\mathbf{g} \in \mathbb{R}^n$ is the vector of the evaluation of the data-term g on the grid $\Omega \times \Gamma$:

$$g(\gamma, i) = \sum_{j\in\Omega} w_{i,j} |\gamma - v(i)|^p, \qquad i \in \Omega, \gamma \in \Gamma.$$

The practical implementation however requires a lot of memory if the image size $|\Omega|$ and the gray level grid $|\Gamma|$ are large.

4 Experimental analysis

In this section, several experiments are performed to analyze the influence of the different parameters and weights proposed in Section 2. We first study in Section 4.1 the influence of the filtering parameters h and λ in the different models with $p = 1$. Section 4.2 is devoted to the choice of the weights in E_{RNL1}. Finally, Section 4.3 illustrates the influence of the power p in the RNLp data term. All these experiments are carried out on two subimages of *Simpson* and *Barbara* (see Figure 3 for these subimages and Figure 2 for the complete images). In this section, the noise level is always set to $\rho = 30\%$. Results are provided both under the form of PSNR tables and restored images. Recall that the PSNR is a way to measure the quality of a restored image $\hat{u}$ in comparison to an uncorrupted one u^0. It is given by the formula

$$\mathrm{PSNR}(u^0, \hat{u}) = 10 \log_{10} \frac{\gamma_1^2 |\Omega|}{\sum_{i\in\Omega} (u^0(i) - \hat{u}(i))^2},$$

where $|\Omega|$ is the size of the support of u^0 and γ_1 the maximum intensity of u^0 (typically, $\gamma_1 = 255$). More systematic denoising results for different complete images and different noise levels are provided in Section 5.

4.1 Influence of the parameters h and λ

Figures 3, 4 and 5 show the denoising results obtained by minimizing E_{RNL1} with exponential weights (given by Equation (5)) for different values of the parameters h and λ. Recall that for a given value of h, the parameter λ controls the amount of smoothing of the total variation term. The influence of h is more subtle since it both controls the number of similar patches taken into account in the restoration and the balance between the Non-Local-Median and the TV regularization.

Note that for a given value of h, when λ tends toward 0, the energy E_{RNL1} converges to E_{NL1}. On the contrary, when λ is fixed and h decreases toward 0, then E_{RNL1} with (normalized or not) exponential weights tends toward $E_{\text{TV-L1}}$. Both models can thus be seen as limits of E_{RNL1} for particular values of λ and h.

Fig. 2 Images used in the experiments. From top to bottom, left to right : *Baboon, Barbara, Boat, Bois, Bridge, Cameraman, Converse, Goldhill, Lena, Peppers, SanFrancisco* and *Simpson*.

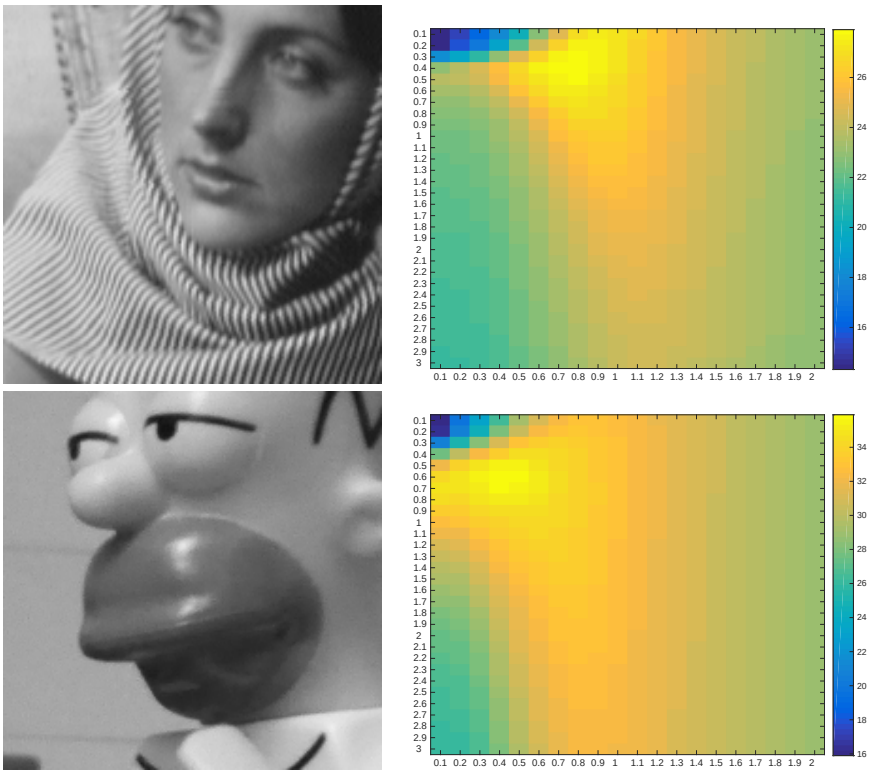

Fig. 3 Influence of h and λ on the results of E_{RNL1}. On each line, the original extract image is shown the left, and the PSNR results are on the right. The vertical axis of the table represents the values of $\lambda \in [0.1 : 0.1 : 3]$ and the horizontal axis represents the values of $h \in [0.1 : 0.1 : 2]$. The impulse noise intensity is $\rho = 0.3$.

On Figures 4 and 5, we observe that for $\lambda = 0$ and for small values of h, residual impulse dots are still present in the denoised images. For these small values of the filtering parameter h, the number of patches that

are taken into account in the Non-Local Median is too small at some points, particularly in rare or textured areas. This can be corrected by choosing a larger value for h. This is indeed necessary to restore properly the texture on *Barbara*, but at the cost of oversmoothing the details. This is less important on the much more regular image *Simpson*. As a consequence, the optimal value for the parameter h is larger on *Barbara*.

Another solution to remove the remaining impulse dots is to increase λ, which usually allows to smooth out the residual impulse noise not discarded by the Non-Local-Median. It is particularly useful for regular regions, like those of *Simpson*, but results in very poor texture recovery. Observe on Figure 3 that the value of λ giving the best performance is usually smaller for textures than for smooth regions, illustrating the more contrasted behavior of the TV regularization on such areas.

4.2 Choosing the weights

We now discuss the choice of judicious weights in the RNL1 model. We consider all the weights defined in Section 2:

- the weights w^δ, in which case the model boils down to TV-L1,
- the exponential weights w^P between patches,
- the normalized exponential weights $\widetilde{w}^P$,
- the weights w^{NN} that take only into account the nearest neighbors of each patch.

4.2.1 The weights w^δ

The energy E_{RNL1} with weights w^δ is equal to $E_{\text{TV-L1}}$. As explained before, this energy can also be seen as a limit of E_{RNL1} with exponential weights w^P when the parameter h tends toward 0. The results are thus very similar to those displayed on the first columns of Figures 4 and 5 obtained with $h = 0.1$. The TV-L1 model is unable to handle textures properly and has a tendency to turn them into molten areas, resulting in very poor PSNR results on such regions. On the contrary, the model is much better suited for a regular image such as *Simpson*.

4.2.2 Influence of normalization: w^P or $\widetilde{w}^P$?

Figure 6 illustrates the influence of normalizing the exponential weights in the energy. With normalized weights, the regularizing power of λ becomes similar everywhere, since it is balanced by a nonlocal term whose total weight is always one. On the contrary, with unnormalized weights, λ will have a stronger influence on singular regions with few similar patches. On an image of uniform regularity, like *Simpson*, both models seems to provide similar results, probably because the sum of weights at each pixel is close to a constant. On a more complex image like *Barbara*, normalizing the weights reduces the denoising performance of the model.

This is also illustrated on Figure 7, which shows the optimal results of RNL1 on the extract of *Barbara* with 30% of impulse noise. On the left, the optimal result with unnormalized weights is obtained for $h = 0.8$ and $\lambda = 0.4$. In the middle, the optimal result with normalized weights is obtained with $h = 0.8$ and $\lambda = 0.1$. Observe that both results are quite similar on regular areas but that some impulse dots are still present when the weights are normalized.

4.2.3 Exponential weights or fixed number of neighbors?

Our last experiment in this section compares exponential weights with binary weights taking into account a fixed number of neighbors for each image patch. The corresponding PSNR tables are shown on Figure 8. Note that when the number of neighbors is fixed for each point, the relative weights of the data term and of the smoothing term are identical everywhere. The quantitative results obtained on these subimages with a fixed number of neighbors are similar or slightly better than those obtained with exponential weights. However, we observed that in practice, fixing the same number of neighbors for each pixel results in the same artifact as normalizing the weights (see Figure 7, right).

4.3 Influence of the power p in RNLp

As advocated in [18], the choice $p = 0$ provides good results in the NLp-Regression setting, in particular for strong impulse noise. Together with optimization considerations, this motivates the choice for a rather nonconvex though continuous model with $0 < p < 1$. Figures 9 and 10 illustrate the performance of the RNLp model for different values of p on the two test images corrupted by impulse noise with parameter $\rho = 0.3$ and 0.5 respectively. For a low noise level, all methods efficiently remove noise. The *Simpson* image benefits from good results with all three settings, and *Barbara* has better preserved textures on the scarf with $p < 1$, while with a lower p some residual noise remains on the fine textures. The same observations seem to apply to a higher noise level. The nonconvex model with $p < 1$ offers slightly better performance, in particular

Fig. 4 Influence of λ and h on the results of E_{RNL1}, with an impulse noise intensity $\rho = 0.3$. On this figure, we show the effect of varying the smoothing parameter λ and the parameter h used in the exponential weights. It can be seen that these two parameters have a significant impact on the denoising results. In the middle, we show the result (and the corresponding values of h and λ) that has the best PSNR.

for the smooth image *Simpson*. However for *Barbara*, more residual noise on the scarf can be observed for $p = 0.2$. All in all, a power $p = 0.5$ seems to offer overall good performance, as it acts as a compromise between the nonconvex model which offers good behavior for strong impulse noise, while behaving more smoothly. Note also that for smaller values of p, the nonconvexity gets more important, hence convergence is harder to achieve.

5 Denoising results

5.1 RNL1

This section gathers experiments comparing RNL1 with the classical models TV-L1 and NL-Median. We also provide a comparison with the recent algorithm PARIGI [18]. Tables 2 and 3 illustrate the PSNR performance of the different models for several images, displayed on Figure 2. All images have the same size (512×512), with the exception of *Simpson*, whose size is 1024×1024. All results of the models RNL1, TV-L1 and NL-Median are obtained by optimizing the param-

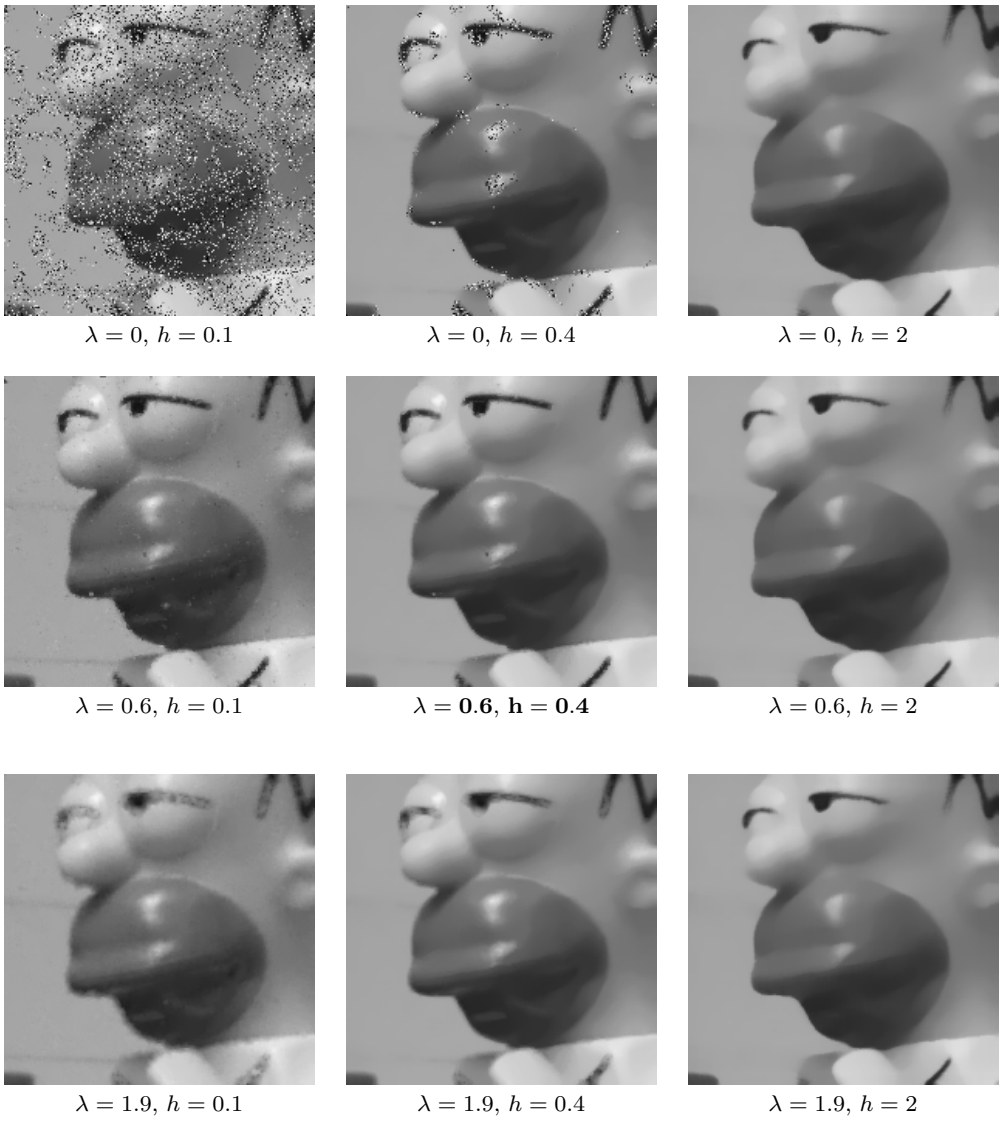

Fig. 5 Influence of λ and h on the results of E_{RNL1}, with an impulse noise intensity $\rho = 0.3$. On this figure, we show the effect of varying the smoothing parameter λ and the parameter h used in the exponential weights. It can be seen that these two parameters have a significant impact on the denoising results. In the middle, we show the result (and the corresponding values of h and λ) that has the best PSNR.

eters λ and h. Figures 11 and 12 show a zoom on the results of the different algorithms on *Simpson*, *Baboon*, *Barbara* and *Bois* when $\rho = 20\%$ and $\rho = 40\%$.

First, observe that the results of TV-L1 and NL-Median are complementary. The variational model gives excellent results on smooth images, while the non-local model performs much better on textured images. This is especially visible on an image like *Barbara* (see Figures 11 and 12) where the TV-L1 model is clearly not able to handle the texture properly, while the Non-Local Median restores them almost perfectly. At the same time, the Non-Local Median shows a tendency to oversmooth the details and fine regions. Most of the time, TV-L1 gives better results for small amounts of noise, but its performance decreases faster when the noise level ρ increases. The gap between the two models also tends to decrease with the level of noise, and they appear to be more or less equivalent (PSNR-wise) for most images when $\rho = 50\%$. The RNL1 model can be seen as a generalization of both TV-L1 and NL-Median, since it boils down to the NL-Median when $\lambda = 0$ and is equivalent to TV-L1 when the parameter h tends toward 0. It follows that good choices of these two parameters always permit to reach better restoration results with RNL1 than with those two models. Observe on Figures 11 and 12 how RNL1 is able to

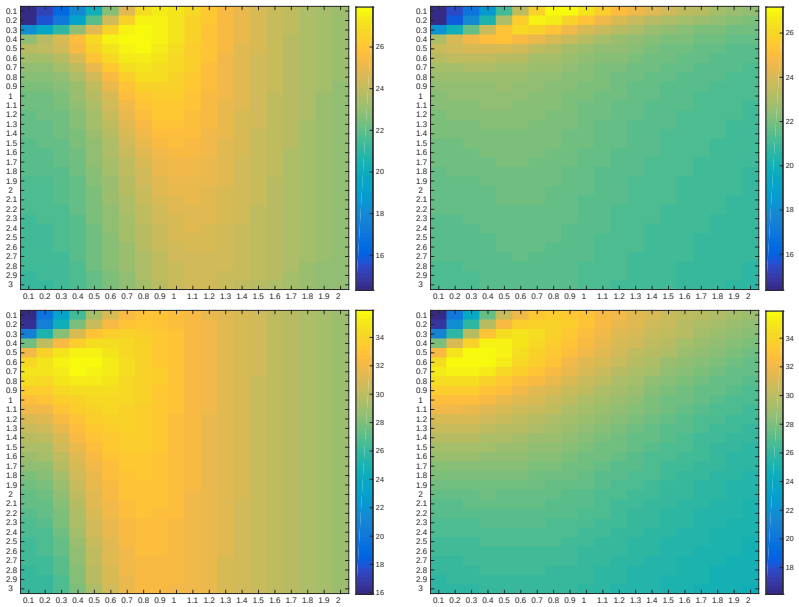

Fig. 6 Influence of normalization. Each figure displays the PSNR as a function of $\lambda \in [0.1 : 0.1 : 3]$ on the vertical axis and of $h \in [0.1 : 0.1 : 2]$ on the horizontal axis. On the first line, extract of *Barbara* and extract of *Simpson* on the second line. Left: unnormalized weights. Right: normalized weights.

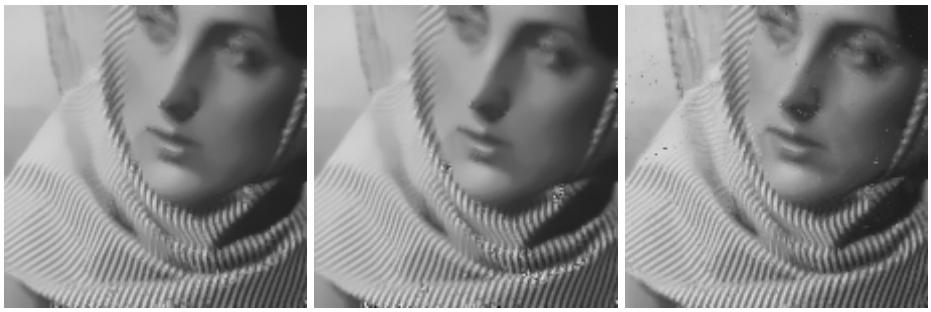

Fig. 7 Results of RNL1 on the extract of *Barbara* with 30% of impulse noise. Left, with unnormalized weights, $h = 0.8$ and $\lambda = 0.4$. Middle, with normalized weights, $h = 0.8$ and $\lambda = 0.1$. Right, with a fixed number of $n = 11$ of neighbors and $\lambda = 0.7$. All parameters are chosen to optimize the PSNR. Observe that the three results are quite similar on regular areas but that some impulse dots are still present on the middle and on the right images.

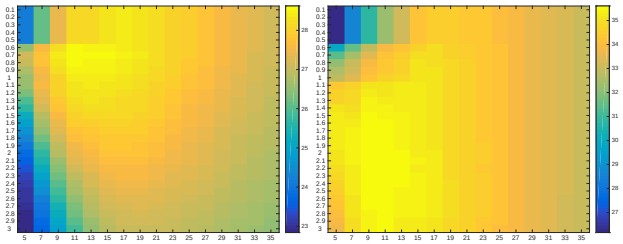

Fig. 8 Same experiment with a fixed number of nearest neighbors instead of exponential weights. The x-axis represents the number of nearest neighbors in the range [5:35]. On the left, extract of *Barbara*. On the right, extract of *Simpson*.

handle textures properly while preserving much more image details than the NL-Median.

For most images, the PSNR results of RNL1 are on par with the results of the recent PARIGI algorithm [18], even if the restored images are visually quite different. PARIGI can be seen as a refinement of the NL-Median. It is also purely patch-based, without additional regularization, but instead of a nonlocal median it computes a Maximum Likelihood at each pixel and happens to be more robust for large values of ρ. In practice, this algorithm proves to be very effective and far better than other approaches on very textured

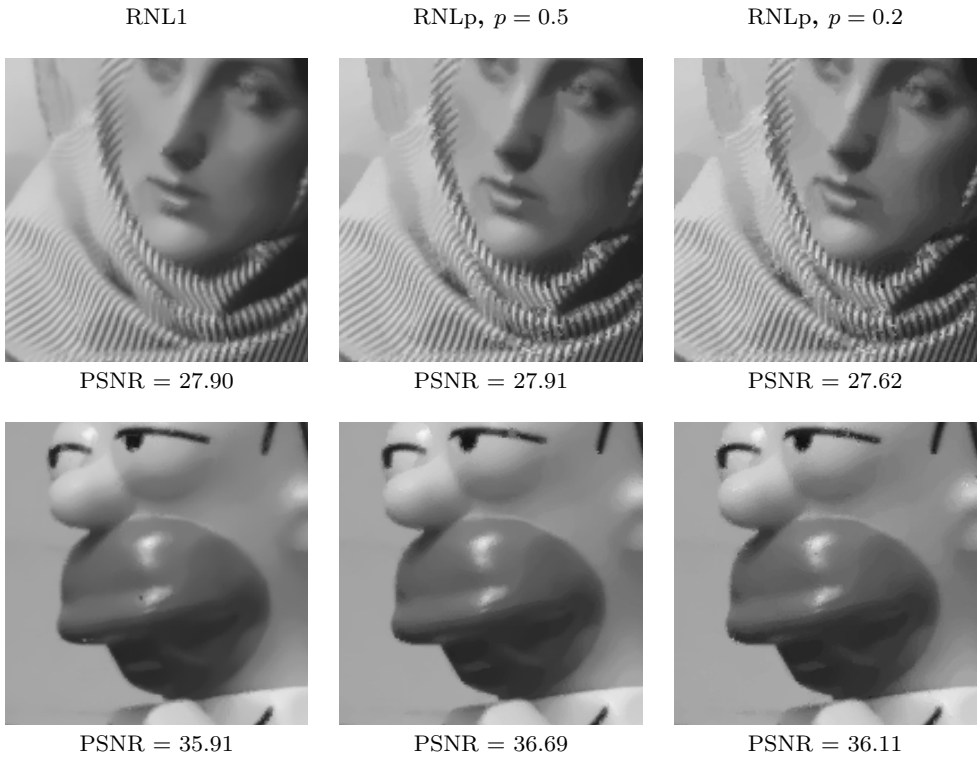

Fig. 9 Denoising of *Barbara* and *Simpson* corrupted with 30% of impulse noise, using the RNLp model with $p = 1$ (left), $p = 0.5$ (middle) and $p = 0.2$ (right).

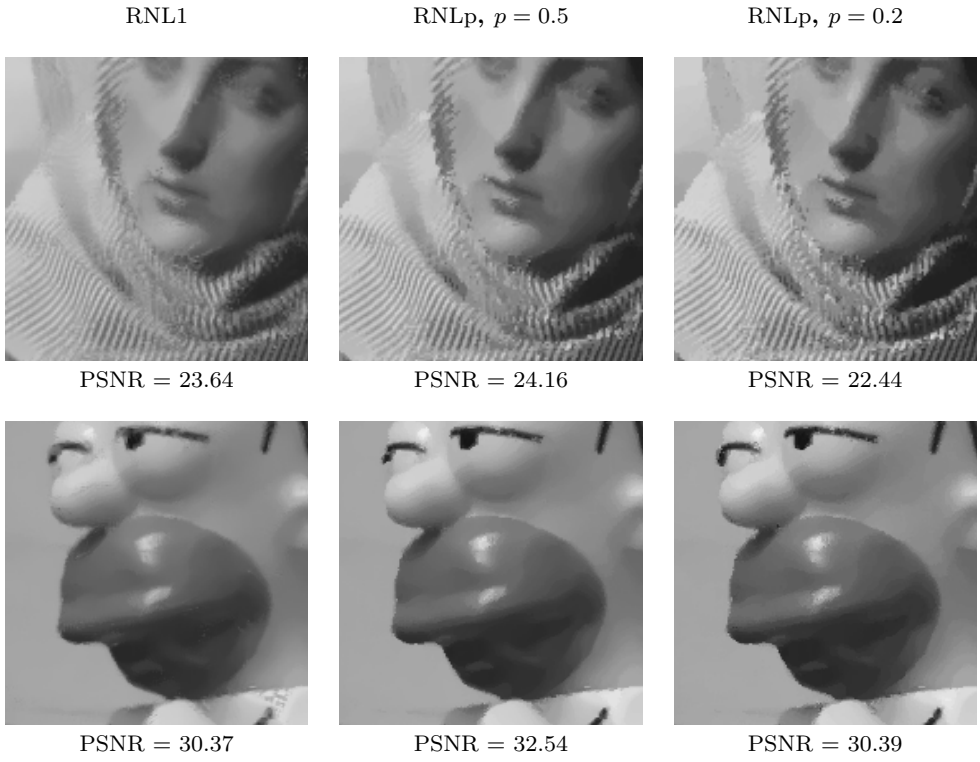

Fig. 10 Denoising of *Barbara* and *Simpson* corrupted with 50% of impulse noise, using the RNLp model with $p = 1$ (left), $p = 0.5$ (middle) and $p = 0.2$ (right).

images like *Barbara*. However, if it performs very well quantitatively, we can see on Figures 11 and 12 that its restoration results tend to present the same artifacts (loss of details, oversmoothing) as the NL-Median, even if to a lesser extent. In practice, we think that images restored with RNL1 present a better balance between texture preservation, restoration of smooth regions and detail recovery, especially for small values of ρ.

A possible refinement proposed in the PARIGI algorithm (but not used in the results of Table 2 and Table 3) consists in applying the algorithm twice after mixing the original noisy image and the restoration result thanks to a map of detected noisy pixels. Such a refinement could also be considered for RNL1. Observe also that none of the models studied in this paper make use of an explicit noise detection (such as the ROAD or ROLD detectors). The detection is only weakly contained in the robust distance presented in Section 2. It goes without saying that taking into account the result of external noise detectors could in practice improve the different denoising results. However, let us point out here that the goal of this paper is not to yield the best possible denoising results but rather to thoroughly compare and understand the advantages and weaknesses of different models in presence of impulse noise.

5.2 RNLp

Due to the heavy computations involving the nonconvex optimization when $p < 1$, we have devoted the comparisons to a selected number of images. Note that the performance of the method for $p < 1$ depends on the discretization of the gray level set Γ, i.e., on the number of gray levels chosen to represent the level sets. In order to limit the amount of required memory and computation time, we have chosen to consider only 128 gray levels for the full size images and 64 for the larger *Simpson* image (i.e. a subsampling factor $\kappa = 2$ and 4 respectively, see Section 3.2). The results could be improved with a finer intensity range. Figures 11 and 12 include the result obtained with RNLp, with $p = 0.5$. The PSNR results for these images are also compared to the above studied methods in Table 1. These results show that the RNLp model is on par with RNL1.

6 Conclusion

In this paper, we have shown that a hybrid model composed of a Non-Local Median data term and a TV regularization could be a solid basis for impulse noise reduction. Other hybrid models had been studied in the literature, especially with nonlocal regularization term. However, to the best of our knowledge, this is the first study making use of a nonlocal L^p data term, with $p \leq 1$, adapting the recent work of Sutour *et al.* [52]. We have shown how to minimize this nonsmooth model in practice, both in the convex ($p = 1$) and nonconvex ($p < 1$) cases. It is illustrated on several experiments that this approach permits to attain state of the art denoising performance for different type of images and different levels of noise. The proposed model, using adaptive weights, is generic enough to include external additive information on the image to be restored or on the noise, such as a precomputed noise map for instance.

References

1. Alliney, S.: Digital filters as absolute norm regularizers. IEEE Trans. Signal Process. **40(6):1548–1562** (1992)
2. Awate, S., Whitaker, R.: Unsupervised, information-theoretic, adaptive image filtering for image restoration. Pattern Analysis and Machine Intelligence, IEEE Transactions on **28**(3), 364–376 (2006)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning **3(1):1–122** (2011)
4. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2, pp. 60–65. IEEE (2005)
5. Candes, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted l1 minimization. Journal of Fourier analysis and applications **14**(5), 877–905 (2008)
6. Chambolle, A., Cremers, D., Pock, T.: A convex approach to minimal partitions. SIAM Journal on Imaging Sciences **5**(4), 1113–1158 (2012)
7. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2011)
8. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numerica **25**, 161–319 (2016)
9. Chan, R., Hu, C., Nikolova, M.: An iterative procedure for removing random-valued impulse noise. Signal Processing Letters, IEEE **11**(12), 921–924 (2004)
10. Chaudhury, K.N., Singer, A.: Non-local Euclidean medians. Signal Processing Letters, IEEE **19**(11), 745–748 (2012)
11. Chen, S., Donoho, D., Saunders, M.: Atomic decomposition by basis pursuit. SIAM Journal Sci. Comput. **20(1):33–61** (1998)
12. Chen, T., Wu, H.: Adaptive impulse detection using center-weighted median filters. Signal Processing Letters, IEEE **8**(1), 1–3 (2001)
13. Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Fixed-point algorithms for inverse problems in science and engineering, pp. 185–212. Springer (2011)

Fig. 11 Zooms of comparative results on *Simpson*, *Baboon*, *Barbara* and *Bois* with $\rho = 20\%$ of random-valued impulse noise. From top to bottom: original images, noisy images, minimizers of $E_{\text{TV-L1}}$, of E_{RNL1}, of E_{RNLp} ($p = 0.5$), of E_{NL1} and the results of PARIGI. Images should be seen at full resolution on the electronic version of the paper.

Fig. 12 Zooms on comparative results on *Simpson*, *Baboon*, *Barbara* and *Bois* with $\rho = 40\%$ of random-valued impulse noise. From top to bottom: original images, noisy images, minimizers of $E_{\text{TV-L1}}$, of E_{RNL1}, of E_{RNLp} ($p = 0.5$), of E_{NL1} and the results of PARIGI. Images should be seen at full resolution on the electronic version of the paper.

		TV-L1	NL-Median	PARIGI (1 it.)	RNL1	RNLp, $p = 0.5$
Baboon	$\rho = 20\%$	24.17	22.89	25.04	24.96	**25.05**
	$\rho = 40\%$	21.56	21.42	21.96	**22.24**	22.01
Barbara	$\rho = 20\%$	25.71	29.83	**34.16**	30.89	30.57
	$\rho = 40\%$	23.48	27.13	**29.47**	27.49	27.67
Bois	$\rho = 20\%$	23.99	22.69	23.19	24.82	**24.88**
	$\rho = 40\%$	21.40	21.41	21.32	22.25	**22.32**
Simpson	$\rho = 20\%$	37.85	35.98	39.55	**39.65**	38.92
	$\rho = 40\%$	33.08	33.25	35.83	35.73	**36.23**

Table 1 PSNR results of different restoration methods for the 512×512 images *Baboon*, *Barbara* and *Bois*, and for the 1024×1024 image *Simpson*. All results of RNLp, RNL1, TV-L1 and NL-Median are obtained by optimizing the parameters λ and h.

		TV-L1	NL-Median	PARIGI (1 it.)	RNL1
Baboon	$\rho = 10\%$	26.59	24.01	26.92	**27.51**
	$\rho = 20\%$	24.17	22.89	**25.04**	24.96
	$\rho = 30\%$	22.64	22.18	23.40	**23.44**
	$\rho = 40\%$	21.56	21.42	21.96	**22.24**
	$\rho = 50\%$	20.69	20.55	20.72	**21.09**
Barbara	$\rho = 10\%$	28.59	31.03	**36.84**	32.67
	$\rho = 20\%$	25.71	29.83	**34.16**	30.89
	$\rho = 30\%$	24.45	28.65	**31.75**	29.25
	$\rho = 40\%$	23.48	27.13	**29.47**	27.49
	$\rho = 50\%$	22.54	24.97	**27.07**	25.23
Boat	$\rho = 10\%$	32.00	28.46	**34.40**	34.05
	$\rho = 20\%$	29.61	27.64	**31.97**	31.33
	$\rho = 30\%$	27.99	26.89	**29.84**	29.55
	$\rho = 40\%$	26.28	25.84	27.62	**27.81**
	$\rho = 50\%$	24.61	24.55	25.44	**25.85**
Bois	$\rho = 10\%$	26.31	23.63	24.50	**27.24**
	$\rho = 20\%$	23.99	22.69	23.19	**24.82**
	$\rho = 30\%$	22.63	21.99	22.31	**23.45**
	$\rho = 40\%$	21.40	21.41	21.32	**22.25**
	$\rho = 50\%$	20.20	20.28	20.21	**20.75**
Bridge	$\rho = 10\%$	28.76	25.27	29.33	**30.22**
	$\rho = 20\%$	26.41	24.38	27.25	**27.61**
	$\rho = 30\%$	24.79	23.80	25.70	**25.93**
	$\rho = 40\%$	23.28	22.81	24.07	**24.27**
	$\rho = 50\%$	21.77	21.54	**22.55**	**22.56**
Cameraman	$\rho = 10\%$	35.62	31.60	37.48	**38.02**
	$\rho = 20\%$	32.82	30.80	34.60	**35.19**
	$\rho = 30\%$	30.36	29.60	32.03	**32.75**
	$\rho = 40\%$	28.11	28.29	29.31	**30.44**
	$\rho = 50\%$	25.04	25.92	26.45	**27.30**

Table 2 PSNR results of different restoration methods for the 512×512 images *Baboon*, *Barbara*, *Boat*, *Bois*, *Bridge* and *Cameraman*. All results of RNL1, TV-L1 and NL-Median are obtained by optimizing the parameters λ and h.

14. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optimization Theory and Applications **158**(2), 460–479 (2013)
15. Condat, L.: Discrete total variation: New definition and minimization. SIAM Journal on Imaging Sciences **10**(3), 1258–1290 (2017)
16. Deledalle, C.A., Denis, L., Tupin, F.: Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. IEEE Trans. Image Process. **18(12):2661–2672** (2009)
17. Deledalle, C.A., Tupin, F., Denis, L.: Poisson NL means: Unsupervised non local means for Poisson noise. In: 2010 IEEE Int. Conf. Image Process. (ICIP). IEEE Signal Process Soc (2010)
18. Delon, J., Desolneux, A.: A patch-based approach for removing mixed Gaussian-impulse noise. SIAM Journal on Imaging Sciences **6**(2), 1140–1174 (2013). http://dx.doi.org/10.1137/120885000
19. Delon, J., Desolneux, A., Guillemot, T.: PARIGI: a Patch-based Approach to Remove Impulse-Gaussian noise from Images. Image Processing On Line **6**, 130–154 (2016). DOI 10.5201/ipol.2016.161
20. Dong, Y., Chan, R., Xu, S.: A detection statistic for random-valued impulse noise. Image Processing, IEEE Transactions on **16**(4), 1112–1120 (2007)
21. Dong, Y., Xu, S.: A New Directional Weighted Median Filter for Removal of Random-Valued Impulse Noise. IEEE Signal Processing Letters **14**(3), 193–

		TV-L1	NL-Median	PARIGI (1 it.)	RNL1
Converse	$\rho = 10\%$	25.94	25.22	25.73	**28.73**
	$\rho = 20\%$	23.03	23.37	24.04	**26.01**
	$\rho = 30\%$	21.40	22.00	22.16	**24.02**
	$\rho = 40\%$	19.99	20.64	20.51	**22.06**
	$\rho = 50\%$	18.28	18.77	18.52	**19.68**
Goldhill	$\rho = 10\%$	33.24	29.72	32.89	**35.53**
	$\rho = 20\%$	30.84	29.07	31.76	**32.76**
	$\rho = 30\%$	29.51	28.6	30.81	**31.12**
	$\rho = 40\%$	27.94	27.78	**29.50**	29.45
	$\rho = 50\%$	26.33	26.35	**27.91**	27.55
Lena	$\rho = 10\%$	36.63	32.25	35.46	**37.26**
	$\rho = 20\%$	33.10	31.43	34.18	**34.91**
	$\rho = 30\%$	30.96	30.49	32.60	**32.85**
	$\rho = 40\%$	29.12	29.34	**31.01**	**31.04**
	$\rho = 50\%$	27.27	27.80	**28.98**	**29.03**
Peppers	$\rho = 10\%$	34.38	32.12	34.97	**36.01**
	$\rho = 20\%$	31.95	30.00	**33.68**	33.38
	$\rho = 30\%$	30.25	28.98	**32.55**	31.26
	$\rho = 40\%$	28.26	27.76	**30.77**	29.28
	$\rho = 50\%$	25.86	26.32	**28.77**	27.20
SanFrancisco	$\rho = 10\%$	25.96	24.99	25.42	**28.71**
	$\rho = 20\%$	23.11	23.26	23.70	**25.94**
	$\rho = 30\%$	21.61	22.18	22.66	**24.14**
	$\rho = 40\%$	20.17	20.74	20.76	**22.20**
	$\rho = 50\%$	18.80	19.19	18.94	**20.38**
Simpson	$\rho = 10\%$	40.90	36.93	41.48	**41.83**
	$\rho = 20\%$	37.85	35.98	39.55	**39.65**
	$\rho = 30\%$	35.62	35.00	**37.90**	**37.93**
	$\rho = 40\%$	33.08	33.25	**35.83**	35.73
	$\rho = 50\%$	30.17	30.85	**33.36**	32.34

Table 3 PSNR results of different restoration methods for the 512×512 images *Converse*, *Goldhill*, *Lena*, *Peppers*, *SanFrancisco* and the 1024×1024 image *Simpson*. All results of RNL1, TV-L1 and NL-Median are obtained by optimizing the parameters λ and h.

196 (2007). DOI 10.1109/LSP.2006.884014. URL http://dx.doi.org/10.1109/LSP.2006.884014
22. Duval, V.: Variational and non-local methods in image processing: a geometric study. Ph.D. thesis, Télécom ParisTech (2011)
23. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association **96**(456), 1348–1360 (2001)
24. Foucart, S., Lai, M.J.: Sparsest solutions of underdetermined linear systems via lq-minimization for 0¡ q ¡1. Applied and Computational Harmonic Analysis **26**(3), 395–407 (2009)
25. Garnett, R., Huegerich, T., Chui, C., He, W.: A universal noise removal algorithm with an impulse detector. Image Processing, IEEE Transactions on **14**(11), 1747–1754 (2005)
26. Gilboa, G., Osher, S.: Nonlocal linear image regularization and supervised segmentation. Multiscale Modeling & Simulation **6**(2), 595–630 (2007)
27. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Modeling & Simulation **7**(3), 1005–1028 (2008)
28. Glowinski, R., Marroco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique **9(R2):41–76** (1975)
29. Goldstein, T., Osher, S.: The split bregman method for l1-regularized problems. SIAM Journal Imaging Sci. **2(2):323–343** (2009)
30. Holler, M., Kunisch, K.: On infimal convolution of tv-type functionals and applications to video and image reconstruction. SIAM Journal on Imaging Sciences **7**(4), 2258–2300 (2014)
31. Hu, H., Li, B., Liu, Q.: Removing mixture of gaussian and impulse noise by patch-based weighted means. Journal of Scientific Computing **67**(1), 103–129 (2016). DOI 10.1007/s10915-015-0073-9. URL http://dx.doi.org/10.1007/s10915-015-0073-9
32. Huang, T., Dong, W., Xie, X., Shi, G., Bai, X.: Mixed noise removal via laplacian scale mixture modeling and nonlocal low-rank approximation. IEEE Transactions on Image Processing **26**(7), 3171–3186 (2017)
33. Huang, Y., Ng, M., Wen, Y.: Fast image restoration methods for impulse and Gaussian noises removal. Signal Processing Letters, IEEE **16**(6), 457–460 (2009)
34. Ishikawa, H.: Exact optimization for markov random fields with convex priors. IEEE transactions on pattern analysis and machine intelligence **25**(10), 1333–1336 (2003)
35. Kervrann, C., Boulanger, J.: Optimal spatial adaptation for patch-based image denoising. IEEE Transactions on Image Processing **15**(10), 2866–2878 (2006)
36. Kervrann, C., Boulanger, J., Coupé, P.: Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In: Scale Space and Varia-

tional Methods in Computer Vision (SSVM), pp. 520–532. Springer (2007)
37. Ko, S.J., Lee, Y.H.: Center weighted median filters and their applications to image enhancement. IEEE Transactions on Circuits and Systems, 38(9):984–993 (1991)
38. Kolmogorov, V., Zabin, R.: What energy functions can be minimized via graph cuts? IEEE transactions on pattern analysis and machine intelligence **26**(2), 147–159 (2004)
39. Lebrun, M., Buades, A., Morel, J.M.: A Nonlocal Bayesian Image Denoising Algorithm. SIAM J. Imaging Sci. **6**(3), 1665–1688 (2013). DOI 10.1137/120874989. URL http://epubs.siam.org/doi/abs/10.1137/120874989
40. Lebrun, M., Colom, M., Buades, A., Morel, J.M.: Secrets of image denoising cuisine. Acta Numerica **21**, 475–576 (2012)
41. Li, B., Liu, Q., Xu, J., Luo, X.: A new method for removing mixed noises. Science China Information sciences **54**(1), 51–59 (2011)
42. Li, Y., Osher, S.: A new median formula with applications to PDE based denoising. Communications in Mathematical Sciences **7**(3), 741–753 (2009)
43. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process. **41(12):3397–3415** (1993)
44. Motta, G., Ordentlich, E., Ramírez, I., Seroussi, G., Weinberger, M.J.: The dude framework for continuous tone image denoising. In: IEEE International Conference on Image Processing (ICIP) 2005, vol. 3, pp. III–345. IEEE (2005)
45. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM journal on computing **24**(2), 227–234 (1995)
46. Nikolova, M.: Minimizers of Cost-Functions Involving Nonsmooth Data-Fidelity Terms. Application to the Processing of Outliers. SIAM J. Numer. Anal. **40**(3), 965–994 (2002). DOI 10.1137/s0036142901389165. URL http://dx.doi.org/10.1137/s0036142901389165
47. Nikolova, M.: A variational approach to remove outliers and impulse noise. Journal of Mathematical Imaging and Vision **20**(1), 99–120 (2004)
48. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: Computer Vision (ICCV), 2011 IEEE International Conference on, pp. 1762–1769. IEEE (2011)
49. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. SIAM Journal on Imaging Sciences **3**(4), 1122–1145 (2010)
50. Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A convex formulation of continuous multi-label problems. Computer Vision–ECCV 2008 pp. 792–805 (2008)
51. Pratt, W.K.: Median filtering. Technical report, Image Proc. Inst., Univ. Southern California (1975)
52. Sutour, C., Deledalle, C.A., Aujol, J.F.: Adaptive regularization of the NL-means: Application to image and video denoising. IEEE Trans. Image Process. **23(8):3506–3521** (2014)
53. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288 (1996)
54. Wang, Y.Q., Morel, J.M.: SURE Guided Gaussian Mixture Image Denoising. SIAM J. Imaging Sci. **6**(2), 999–1034 (2013). DOI 10.1137/120901131. URL http://epubs.siam.org/doi/abs/10.1137/120901131
55. Xiao, Y., Zeng, T., Yu, J., Ng, M.K.: Restoration of images corrupted by mixed Gaussian-impulse noise via l_1-l_0 minimization. Pattern Recognition **44**(8), 1708–1720 (2011). http://dx.doi.org/10.1016/j.patcog.2011.02.002
56. Xiong, B., Yin, Z.: A universal denoising framework with a new impulse detector and nonlocal means. Image Processing, IEEE Transactions on **21**(4), 1663–1675 (2012). URL http://dx.doi.org/10.1109/TIP.2011.2172804
57. Yan, M.: Restoration of images corrupted by impulse noise and mixed gaussian impulse noise using blind inpainting. SIAM Journal on Imaging Sciences **6**(3), 1227–1245 (2013)
58. Yu, G., Sapiro, G., Mallat, S.: Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. IEEE Trans. Image Process. **21**(5), 2481–99 (2012). DOI 10.1109/TIP.2011.2176743
59. Yuan, G., Ghanem, B.: l0tv: A new method for image restoration in the presence of impulse noise. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
60. Zhou, Y., Ye, Z., Xiao, Y.: A restoration algorithm for images contaminated by mixed Gaussian plus random-valued impulse noise. Journal of Visual Communication and Image Representation **24**(3), 283–294 (2013)
61. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: 2011 Int. Conf. Comput. Vis., pp. 479–486. IEEE (2011). DOI 10.1109/ICCV.2011.6126278

Appendix

Proof of Proposition 1 :

We will here produce examples where the two energies have different minimizers (when $p \neq 2$), even for different values of λ. We will need the following lemma, whose proof is simple and therefore left to the reader.

Lemma 1 *Let ω and c be strictly positive real numbers. Let $p \geq 0$. Then the minimum of the function f defined on $[0, c]$ by*

$$\forall x \in [0, c], \quad f(x) = |x|^p + \omega |c - x|^p$$

is achieved, when $p > 1$ at

$$x_* = \frac{c\omega^\beta}{1 + \omega^\beta}, \text{ where } \beta = 1/(p-1).$$

When $p \leq 1$, f is concave and the minimum is achieved at

$$x_* = 0 \text{ if } \omega < 1 \text{ and at } x_* = c \text{ if } \omega > 1.$$

When $\omega = 1$ and $p = 1$, f is constant on $[0, c]$, and when $\omega = 1$ and $p < 1$, 0 and c are both minimizers of f.

Let us first consider the case $p > 1$. When $p = 2$, these two energies are equal up to a constant, and have therefore the same minimizers. Now, this is the only

p for which these two energies are equivalent. Indeed, consider the following one dimensional example:

$$1 \le i \le 5, \qquad v = (0,0,1,0,0) \quad \text{and}$$
$$w = (w_{ij})_{1\le i,j\le 5} = \begin{pmatrix} 2/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 2/3 \end{pmatrix}.$$

The $\mathrm{TV}(u)$ term is defined as

$$\mathrm{TV}(u) = \sum_{i=1}^{4} |u(i+1) - u(i)|.$$

We can then compute that u_{NLp} is given by

$$u_{\mathrm{NLp}} = (0, c_p, c_p, c_p, 0),$$
$$\text{where } c_p = \frac{1}{1+2^\beta}, \text{ with } \beta = 1/(p-1).$$

The energy $E_{\text{NLp-TV}}$ is thus

$$E_{\text{NLp-TV}}(u) = |u(1)|^p + \sum_{i=2}^{4} |u(i) - c_p|^p + |u(5)|^p + \lambda\, \mathrm{TV}(u),$$

Whereas the energy E_{RNLp} is

$$E_{\mathrm{RNLp}}(u) = |u(1)|^p + |u(5)|^p + \frac{1}{3}\sum_{i=2}^{4} |u(i) - 1|^p + \frac{2}{3}\sum_{i=2}^{4} |u(i)|^p + \lambda\, \mathrm{TV}(u).$$

For symmetry reasons, $E_{\text{NLp-TV}}$ and E_{RNLp} are both minimized for signals u of the form

$$u = (a, b, b, b, a)$$

with $0 \le a \le b \le 1$, (or c_p for $E_{\text{NLp-TV}}$). We then can write

$$E_{\mathrm{RNLp}}(u) = 2a^p + 2b^p + (1-b)^p + 2\lambda(b-a)$$

and

$$E_{\text{NLp-TV}}(u) = 2a^p + 3(c_p - b)^p + 2\lambda(b-a).$$

When $\lambda > 2p$ then $a \mapsto E_{\mathrm{RNLp}}(u)$ is decreasing and $b \mapsto E_{\mathrm{RNLp}}(u)$ is increasing, therefore E_{RNLp} is minimized for $a = b$. The same result holds for $E_{\text{NLp-TV}}$.

As a consequence, when λ is large, the two energies will both be minimized for constant signals, but the constant is not the same for E_{RNLp} and $E_{\text{NLp-TV}}$. Indeed, the constant signal that minimizes E_{RNLp} has value

$$a = \frac{1}{1+4^\beta}.$$

Whereas the constant signal that minimizes $E_{\text{NLp-TV}}$ has value

$$\widetilde{a} = \frac{c_p 3^\beta}{2^\beta + 3^\beta} = \frac{3^\beta}{(1+2^\beta)(2^\beta+3^\beta)}.$$

The two constants a and $\widetilde{a}$ are different as soon as $p \neq 2$. Indeed we have

$$\begin{aligned} a = \widetilde{a} &\Leftrightarrow 3^\beta(1+4^\beta) = (1+2^\beta)(2^\beta+3^\beta) \\ &\Leftrightarrow 6^\beta = (1+2+3)^\beta = 1+2^\beta+3^\beta \\ &\Leftrightarrow \beta = 1 \Leftrightarrow p = 2. \end{aligned}$$

When $p \le 1$, we consider the same v but different weights w. More precisely, we consider

$$1 \le i \le 5, \qquad v = (0,0,1,0,0) \quad \text{and}$$
$$w = (w_{ij})_{1\le i,j\le 5} = \begin{pmatrix} 1/3 & 0 & 2/3 & 0 & 0 \\ 1/6 & 0 & 2/3 & 0 & 1/6 \\ 0 & 1/3 & 0 & 2/3 & 0 \\ 0 & 0 & 2/3 & 0 & 1/3 \\ 0 & 2/3 & 0 & 1/3 & 0 \end{pmatrix}.$$

We can then compute that u_{NLp} is given, whatever the value of $p \le 1$, by

$$u_{\mathrm{NLp}} = (1,0,1,0,1).$$

And with the same reasoning as above, we have that, when λ is large enough, the minimizers of E_{RNLp} and $E_{\text{NLp-TV}}$ are constants. And a simple computation, using Lemma 1, shows that the minimizer of E_{RNLp} is the constant 0, whereas the minimizer of $E_{\text{NLp-TV}}$ is the constant 1. Notice also that for any $\lambda > 0$, $E_{\mathrm{RNLp}}(0,0,0,0,0) = 2$ and $E_{\mathrm{RNLp}}(1,1,1,1,1) = 3$ while $E_{\text{NLp-TV}}(1,1,1,1,1) = 2$ and $E_{\text{NLp-TV}}(0,0,0,0,0) = 3$, so the minimizer of the first energy will never be a minimizer of the second one, and vice versa, even for a different value of λ.