SIAM J. IMAGING SCIENCES
Vol. 4, No. 2, pp. 651–694

Total Variation as a Local Filter*

Cécile Louchet† and Lionel Moisan‡

Abstract. In the Rudin–Osher–Fatemi (ROF) image denoising model, total variation (TV) is used as a global regularization term. However, as we observe, the local interactions induced by TV do not propagate much at long distances in practice, so that the ROF model is not far from being a local filter. In this paper, we propose building a purely local filter by considering the ROF model in a given neighborhood of each pixel. We show that appropriate weights are required to avoid aliasing-like effects, and we provide an explicit convergence criterion for an associated dual minimization algorithm based on Chambolle's work. We study theoretical properties of the obtained local filter and show that this localization of the ROF model brings an interesting optimization of the bias-variance trade-off, and a strong reduction of an ROF drawback called the "staircasing effect." Finally, we present a new denoising algorithm, TV-means, that efficiently combines the idea of local TV-filtering with the nonlocal means patch-based method.

Key words. total variation, variational model, local filter, image denoising, nonlocal means

AMS subject classifications. 68U10, 65K10, 62H35

DOI. 10.1137/100785855

1. Introduction.

Image denoising/smoothing is one of the most frequently considered issues in image processing, not only because it plays a key preliminary role in many computer vision systems, but also because it is probably the simplest way to address the fundamental issue of image modeling, as a starting point towards more complex tasks such as deblurring, demosaicking, and inpainting. Among denoising/smoothing methods, several classes arise naturally. One of these consists of local filters, that is, translation-invariant operators that transform a gray-level image $v : \Omega \to \mathbb{R}$ into a gray-level image

$$u : x \mapsto T\left((v(x+z))_{z\in B}\right),$$

where B is a bounded set (typically, a disc with radius r centered in 0), and T is an application from $\mathbb{R}^B$ to $\mathbb{R}$. Note that this definition holds equally for images defined on a continuous domain ($\Omega, B \subset \mathbb{R}^2$) or on a discrete domain ($\Omega, B \subset \mathbb{Z}^2$). Local filters include the averaging filter, obtained (in a continuous setting) with the averaging operator

$$T(w) = \frac{1}{|B|}\int_B w(z)\,dz,$$

*Received by the editors February 16, 2010; accepted for publication (in revised form) March 4, 2011; published electronically June 21, 2011.
http://www.siam.org/journals/siims/4-2/78585.html
†MAPMO, CNRS UMR 6628, Université d'Orléans, 45067 Orléans Cedex, France (Cecile.Louchet@univ-orleans.fr).
‡MAP5, CNRS UMR 8145, Université Paris Descartes, 75006 Paris, France (Lionel.Moisan@parisdescartes.fr).

and, more generally, convolution filters with finite impulse response (when T is linear). Let us also mention contrast-invariant operators like, for example, the median filter or the erosion filter associated to

$$T(w) = \inf_{z \in B} w(z).$$

Another important example is given by Yaroslavsky's filter [65], corresponding to

$$T(w) = \frac{1}{C(w)} \int_B w(z)\, e^{-|w(z)-w(0)|^2/h^2}\, dz, \text{ where } C(w) = \int_B e^{-|w(z)-w(0)|^2/h^2}\, dz,$$

and the related (not strictly local) SUSAN [59] and bilateral filters [62] (see [18] for a discussion on the relationship between these three filters).

Another class of denoising/smoothing methods consists of variational formulations, which transform an image v into an image u that minimizes some energy functional $E_\lambda(u)$ depending on v and on a parameter λ. A typical example is the L^2-H^1 minimization associated to

$$E_\lambda(u) = \|u - v\|^2 + \lambda \int_\Omega |\nabla u|^2, \tag{1}$$

where $\nabla u = (\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y})^T$ is the gradient of u. This example is a special case of the Wiener filter, which can be solved explicitly when $\Omega = \mathbb{R}^2$ with

$$\forall \xi \in \mathbb{R}^2, \quad \hat{u}(\xi) = \frac{\hat{v}(\xi)}{1 + \lambda |\xi|^2},$$

where $\hat{f}$ denotes the Fourier transform of $f : \mathbb{R}^2 \to \mathbb{R}$. This filter is a convolution, but since the associated kernel is not compactly supported, it cannot be written as a purely local filter. A more sophisticated example, avoiding undesirable blur effects caused by (1), is obtained with

$$E_\lambda(u) = \|u - v\|^2 + \lambda \int_\Omega |\nabla u|, \tag{2}$$

which corresponds to the Rudin–Osher–Fatemi (ROF) model for image denoising [58]. This model, which is the starting point of this work, has been widely used in image processing for various tasks including denoising [25, 37], deblurring [27, 57], interpolation [42], super-resolution [4, 22], inpainting [28, 26], and cartoon/texture decomposition [10, 53]. Dramatic improvements have also been made recently on accuracy and computation speed for the numerical solving of ROF-derived variational problems [9, 19, 22, 23, 34, 35, 64].

These two classes of methods (not to mention others) correspond to two different points of view. A local filter can be iterated, which generally results in a partial differential equation (PDE) formulation with interesting interpretations (the heat equation for positive isotropic averaging filters, the mean curvature motion for the median filter [6], etc.). The amount of smoothing/denoising can also be increased by changing the size of the neighborhood B, whereas in a variational formulation (energy E_λ), this role is played by the hyperparameter λ that controls the trade-off between the smoothness of u and its distance to the original image v. The use of local filters is very natural in a shape recognition context, where the

possibility of occlusions makes long-distance smoothing interactions questionable (it seems more relevant to smooth the background of a scene mostly independently of the foreground, and vice versa). In general, variational formulations do not lead to local filters, because short-distance interactions involved in $E_\lambda(u)$ (typically resulting from partial derivatives) cause long-distance interactions due to the minimization process [14].

In this work, we first study the importance of these long-distance interactions for the ROF model (2). We show in section 2 that most image pixels have a very limited influence zone around them, which suggests that the ROF model is not far from being a local filter. In section 3, we follow this idea and derive a local filter by considering an ROF model on a neighborhood of each pixel. We show in particular that the introduction of a smooth window (that is, appropriate weights on the neighborhood) is required to avoid aliasing-like artifacts (that is, the enhancement of particular frequencies). The monotony of this local TV-filter is investigated in section 4, and we show that it admits two limiting PDEs: the total variation flow [12] and the heat equation. In section 5, we build a numerical scheme inspired from Chambolle's algorithm [22] to solve the weighted local ROF model on an arbitrary neighborhood, and we give a convergence criterion involving a relationship between the time step and the weighting function. Experiments performed in section 6 show in particular two interesting properties of the local TV-filter compared to the global ROF model: its ability to reach an intermediate bias-variance trade-off for image denoising and the elimination of a well-known artifact called the "staircasing effect." To illustrate the perspectives offered by local TV-filtering, in section 7 we build a new denoising filter, called *TV-means*, that combines in a simple way the strengths of TV-denoising and nonlocal (NL)-means denoising [17], and produces much better results than either of them.

2. How nonlocal is TV-denoising? In this section we investigate the amount of locality of the ROF denoising model. As we shall see, even though TV-denoising requires a global optimization on the whole image, local interactions do not propagate very far, and the gray level of a denoised pixel essentially depends on the pixels lying in its neighborhood, while other pixels have negligible or null impact.

In the following, a (discrete) image is a function $u : \Omega \to \mathbb{R}$, where Ω is a subset of $\mathbb{Z}^2$ (the set of pixels) and $u(i, j)$ represents the gray level at pixel (i, j). If A is a subset of $\mathbb{Z}^2$, A^c will denote its complement, and ∂A is the boundary of A, defined as the set of pixels for which at least one neighbor (for the four-neighbor topology) does not belong to A. To a subset A of $\mathbb{Z}^2$ we associate the set $A^\square \subset \mathbb{R}^2$ defined by

$$A^\square = \bigcup_{(i,j)\in A} \left[i - \frac{1}{2}, i + \frac{1}{2}\right] \times \left[j - \frac{1}{2}, j + \frac{1}{2}\right].$$

Thus, $A^\square$ is obtained by considering grid points of $\mathbb{Z}^2$ as 1×1 squares of $\mathbb{R}^2$. It is interesting to notice that if a discrete image $u : \mathbb{Z}^2 \to \mathbb{R}$ is extended to an image $\bar{u} : \mathbb{R}^2 \to \mathbb{R}$ using the nearest neighbor interpolation, then the level sets of $\bar{u}$ are obtained by applying the $\cdot^\square$ operator to the level sets of u, that is,

$$\forall \lambda \in \mathbb{R}, \qquad \{x \in \mathbb{R}^2,\ \bar{u}(x) \le \lambda\} = \{x \in \mathbb{Z}^2,\ u(x) \le \lambda\}^\square.$$

The $\cdot^{\square}$ operator allows us to extend the usual perimeter operator (defined on Caccioppoli sets of $\mathbb{R}^2$) to discrete sets, with

$$\forall A \subset \mathbb{Z}^2, \quad \text{per}A = \text{Perimeter}(A^{\square}).$$

Let us first recall the principle of ROF denoising [58]. If u is an image defined on $\Omega \subset \mathbb{Z}^2$, its total variation (TV) is defined by

$$TV(u) = \sum_{(i,j)\in\Omega} |\nabla u(i,j)|, \tag{3}$$

where $|\nabla u(i,j)|$ denotes a given scheme of the gradient norm of u at pixel (i,j). If v is a noisy image, the ROF model proposes to smooth it by selecting the unique image $u = T(v)$ that minimizes

$$E_\lambda(u) = \|u - v\|^2 + \lambda TV(u), \tag{4}$$

where $\|\cdot\|$ stands for the classical Euclidean norm on $\mathbb{R}^\Omega$. The positive parameter λ controls the amount of denoising and should be set according to the noise level.

In what follows, as in [22], we shall choose for $\nabla u(i,j)$ a scheme based on simple differences between neighbor pixels, that is,

$$(\nabla u)^1_{i,j} = \begin{cases} u(i+1,j) - u(i,j) & \text{if } (i+1,j) \in \Omega, \\ 0 & \text{else,} \end{cases}$$
$$(\nabla u)^2_{i,j} = \begin{cases} u(i,j+1) - u(i,j) & \text{if } (i,j+1) \in \Omega, \\ 0 & \text{else,} \end{cases} \tag{5}$$

and we shall derive $|\nabla u(i,j)|$ by considering either the ℓ^1-norm or the ℓ^2-norm of $\nabla u(i,j)$ in $\mathbb{R}^2$. Each choice involves its own specificities:

- The ℓ^1-norm of the gradient, given by

$$|\nabla u(i,j)|_1 = |(\nabla u)^1_{i,j}| + |(\nabla u)^2_{i,j}|, \tag{6}$$

is not isotropic (it favors vertical and horizontal directions), but it has the advantage of making TV satisfy the coarea formula [23, 34, 39]

$$\forall u \in \mathbb{R}^\Omega, \quad TV(u) = \int_{\mathbb{R}} \text{per}\{x \in \Omega; u(x) \le \lambda\}\, d\lambda, \tag{7}$$

which allows us to interpret $TV(u)$ as the cumulated length of the level lines of u. Furthermore, ROF denoising in that case is *monotone*, in the sense that

$$[\forall x \in \Omega,\ v_1(x) \le v_2(x)] \implies [\forall x \in \Omega,\ T(v_1)(x) \le T(v_2)(x)] \tag{8}$$

(see [23] for a proof). In this section, these analytic properties will be most useful in studying the locality of TV-denoising; this is why we will consider the ℓ^1-norm.

- The ℓ^2-norm of the gradient, given by

$$|\nabla u(i,j)|_2 = \sqrt{((\nabla u)^1_{i,j})^2 + ((\nabla u)^2_{i,j})^2} \tag{9}$$

is more isotropic, and, in practice, the images denoised with this scheme look slightly more natural. However, neither the coarea formula nor the monotony principle (see section 4) holds anymore. This ℓ^2-norm will be used in the numerical experiments (section 6).

Both ℓ^1- and ℓ^2-schemes are compatible with two other quite basic properties of operator T. Namely, they force the TV-denoising operator to preserve the image average, i.e.,

$$\forall v \in \mathbb{R}^\Omega, \quad \sum_{x\in\Omega} T(v)(x) = \sum_{x\in\Omega} v(x), \tag{10}$$

and to be shift-invariant, i.e.,

$$\forall v \in \mathbb{R}^\Omega,\ \forall b \in \mathbb{R}, \quad T(v+b) = T(v) + b, \tag{11}$$

which will be useful in what follows.

Let us now discuss the locality of TV-denoising, that is, the influence region of an arbitrary pixel in the denoising process. We easily observe that the dependence in E_λ of a gray level $u(x)$ brings into play the pixels y that are neighbors of x only, through their gray levels $u(y)$. But the levels $u(y)$ also depend on their neighbors and so on, such that the denoising is likely to be indeed global, in the sense that a change on a pixel in the noisy image might change the value of any other pixel in the denoised image.

For instance, let Ω be a rectangular domain, and let x_0 be a pixel in $\Omega \setminus \partial\Omega$. If the image v is an impulse image (discrete Dirac) defined for some $A > 0$ by

$$\forall x \in \Omega, \quad v(x) = A \cdot \delta_{x=x_0} = \begin{cases} A & \text{if } x = x_0, \\ 0 & \text{else,} \end{cases} \tag{12}$$

(here δ is the 0-1 Kronecker-delta function), then simple calculations show that $T(v)$, computed with the ℓ^1-scheme, satisfies

$$\forall x \in \Omega, \quad T(v)(x) = \begin{cases} (A - 2\lambda)\delta_{x=x_0} + \frac{\lambda}{\lambda_c} \cdot \frac{A}{|\Omega|}\delta_{x\neq x_0} & \text{if } \lambda < \lambda_c, \\ A/|\Omega| & \text{if } \lambda \geq \lambda_c, \end{cases} \tag{13}$$

where $|\Omega|$ denotes the cardinal of Ω and $\lambda_c = \frac{A}{2} \cdot (1 - \frac{1}{|\Omega|})$. This means that any pixel of the domain can be affected by a change in $v(x_0)$. However, one can notice that when $|\Omega|$ is large, then the change in any pixel $x \neq x_0$ is negligible compared to the change in x_0; hence the globality of TV-denoising is rather weak in this example. Note also that if Ω is finite, the scheme (5) corresponds to Neumann conditions in the minimization of (4), because the gradient at a pixel lying on the boundary $\partial\Omega$ is treated as if its facing neighbors lying outside Ω had the same gray level. TV-denoising based on these Neumann conditions has the property of preserving the image average (10), which explains the slight correction of $\frac{\lambda}{\lambda_c} \cdot \frac{A}{|\Omega|}$ obtained on the pixels $x \neq x_0$ in (13).

2.1. Locality and boundary conditions. In the remainder of section 2, we shall offer evidence of the fact that, with appropriate boundary conditions, TV-denoising is essentially local, both in exact and in simulated examples. We propose here to consider boundary conditions other than Neumann conditions. First we consider *infinite boundary conditions*: an image is continued by 0 on the entire set $\mathbb{Z}^2$ and is viewed as a finitely supported function, that is, a function u whose support

$$\mathrm{supp}(u) = \{x \in \Omega,\, u(x) \neq 0\}$$

is a finite set.

Proposition 1 (locality for an infinite domain). *Let $v \in \ell^2(\mathbb{Z}^2)$. We consider the variational problem*

$$\text{minimize} \sum_{x\in\mathbb{Z}^2} (u(x) - v(x))^2 + \lambda TV(u) \quad \text{for } u \in \ell^2(\mathbb{Z}^2), \tag{14}$$

where $TV(u)$ is computed according to the ℓ^1-scheme (6) with $\Omega = \mathbb{Z}^2$. If v has a finite support, i.e., if there exists a finite rectangle $\Omega_0 \subset \mathbb{Z}^2$ satisfying $\mathrm{supp}(v) \subset \Omega_0$, then the image $u = T(v)$ solving the variational problem (14) is also finitely supported and satisfies $\mathrm{supp}(u) \subset \Omega_0$.

Proof. (i) Let $A \geq 0$ and $v = A\mathbb{1}_{\Omega_0}$, where Ω_0 is any rectangle of $\mathbb{Z}^2$, and where

$$\forall x \in \mathbb{Z}^2, \quad \mathbb{1}_{\Omega_0}(x) = \begin{cases} 1 & \text{if } x \in \Omega_0, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

We prove $\mathrm{supp}(T(v)) \subset \Omega_0$ by showing that $\{x \in \mathbb{Z}^2, T(v)(x) > 0\} \subset \Omega_0$ and then that $T(v)$ is nonnegative. Assume that the upper level set $\{x \in \mathbb{Z}^2, T(v)(x) > 0\}$, abbreviated as $\{T(v) > 0\}$, intersects the complement Ω_0^c of Ω_0. Consider the image $w = T(v) \cdot \mathbb{1}_{\Omega_0}$. Changing $T(v)$ into w decreases the data-fidelity term because

$$\sum_{x\in\mathbb{Z}^2} (w(x) - v(x))^2 = \sum_{x\in\Omega_0} (T(v)(x) - v(x))^2 < \sum_{x\in\mathbb{Z}^2} (T(v)(x) - v(x))^2.$$

It also decreases the TV. Indeed a level set $\{w \geq \lambda\}$ satisfies

$$\{w \geq \lambda\} = \{T(v) \geq \lambda\} \cap \Omega_0,$$

and thanks to the convexity of $\Omega_0^{\square}$ (due to the fact that Ω_0 is a rectangle), we get

$$\begin{aligned} \mathrm{per}\{w \geq \lambda\} &= \mathrm{per}(\{T(v) \geq \lambda\} \cap \Omega_0) \\ &= \mathrm{Perimeter}(\{T(v) \geq \lambda\}^{\square} \cap \Omega_0^{\square}) \leq \mathrm{per}\{T(v) \geq \lambda\}. \end{aligned}$$

Hence thanks to the coarea formula (7), we get

$$TV(w) = \int_{\mathbb{R}} \mathrm{per}\{w \geq \lambda\}\, d\lambda \leq \int_{\mathbb{R}} \mathrm{per}\{T(v) \geq \lambda\}\, d\lambda = TV(T(v)).$$

Finally we get

$$\begin{aligned} E_\lambda(w) &= \sum_{x\in\mathbb{Z}^2} (w(x) - v(x))^2 + \lambda TV(w) \\ &< \sum_{x\in\mathbb{Z}^2} (T(v)(x) - v(x))^2 + \lambda TV(T(v)) = E_\lambda(T(v)), \end{aligned}$$

and we obtain a contradiction to the optimality of $T(v)$.

Now assume that the lower level set $\{T(v) < 0\}$ is nonempty. We consider the positive part $T(v)_+$ of $T(v)$, where $(y)_+ = \max(y, 0)$ for any y. As earlier, we have $\|T(v)_+ - v\|^2 < \|T(v) - v\|^2$ and

$$TV(T(v)_+) = \int_{\mathbb{R}^+} \mathrm{per}(\{T(v) < \lambda\})\, d\lambda \leq \int_{\mathbb{R}} \mathrm{per}(\{T(v) < \lambda\})\, d\lambda = TV(T(v)).$$

This contradicts the optimality of $T(v)$ and proves that $\{T(v) < 0\}$ is empty. The proposition is therefore proved for $v = A\mathbb{1}_{\Omega_0}$.

(ii) Now if v is an arbitrary image satisfying the conditions of the proposition, then

$$-A\mathbb{1}_{\Omega_0} \leq v \leq A\mathbb{1}_{\Omega_0}, \quad \text{where} \quad A = \max_{x \in \Omega_0} |v(x)|.$$

By monotony of TV-denoising (see (8)), the inequalities remain true for the denoised images; i.e.,

$$-T(A\mathbb{1}_{\Omega_0}) \leq T(v) \leq T(A\mathbb{1}_{\Omega_0}),$$

where the leftmost and rightmost terms both have their support included in Ω_0, thanks to part (i) of the proof. Consequently, $\mathrm{supp}(T(v)) \subset \Omega_0$, as claimed. ∎

Now we consider Dirichlet boundary conditions, which impose that the denoised image vanishes on $\partial\Omega$. Then the same kind of result holds, as shown by the following corollary.

Corollary 1 (locality for Dirichlet boundary conditions). *Let Ω_1 be a subset of $\Omega = \mathbb{Z}^2$ and v an image defined on Ω_1. Let us consider the variational problem*

$$\begin{aligned} &\textit{find } u : \mathbb{Z}^2 \to \mathbb{R} \textit{ that minimizes } \sum_{x \in \Omega_1} (u(x) - v(x))^2 + \lambda TV(u) \\ &\textit{subject to} \quad \forall x \in \partial\Omega_1,\ u(x) = 0. \end{aligned} \tag{16}$$

If there exists a rectangle Ω_0 satisfying $\mathrm{supp}(v) \subset \Omega_0 \subset \Omega_1 \setminus \partial\Omega_1$, then the image $u = T(v)$ solving the variational problem (16) satisfies $\mathrm{supp}(u) \subset \Omega_0$.

Proof. As u is defined on $\mathbb{Z}^2$, the implicit Neumann conditions of (5) are discarded, so that the only boundary conditions considered here are the Dirichlet constraints of (16). Let us denote by $\dot{v}$ the image v continued by 0 on the plane $\mathbb{Z}^2$. If u is the solution of (16), then u is also the solution of the following problem: minimize $E_\lambda(u)$ for $u \in \ell^2(\mathbb{Z}^2)$ under the constraint

$$\forall x \in (\mathbb{Z}^2 \setminus \Omega_1) \cup \partial\Omega_1, \quad u(x) = 0, \tag{17}$$

with $E_\lambda(u) = \sum_{x \in \mathbb{Z}^2} (u(x) - \dot{v}(x))^2 + \lambda TV(u)$. Now consider the solution $u' = T(v)$ of problem (14), that is, the minimizer of E_λ over $\ell^2(\mathbb{Z}^2)$. Since $E_\lambda(u') \leq E_\lambda(u)$ and u' satisfies (17) thanks to Proposition 1 (because $v = 0$ on $(\mathbb{Z}^2 \setminus \Omega_1) \cup \partial\Omega_1 \supset \Omega_0^c$), we necessarily have $u' = u$ and consequently $\mathrm{supp}(u) = \mathrm{supp}(u') \subset \Omega_0$. ∎

A consequence of the two results above is that in both cases (infinite domain or Dirichlet boundary conditions), the TV-denoising of the impulse image (12) remains an impulse image, and the pixels apart from x_0 never change.

More generally, if v is finitely supported, then the influence of a pixel $x \in \mathrm{supp}(v)$ is limited to a rectangle Ω_0 (the smallest rectangle containing $\mathrm{supp}(v)$), in the sense that for any $y \in \Omega_0^c$, $T(v)(y)$ does not depend on $v(x)$. This is a first result in favor of the locality of TV-denoising.

2.2. Locality for explicit solutions (continuous domain). The literature about exact solutions of TV-denoising in its continuous version (i.e., when Ω is an open subset of $\mathbb{R}^2$) provides valuable examples of locality in TV-denoising. First, Strong and Chan [60] provide exact solutions to the problem stated in one dimension and also in two dimensions for radially symmetric images, for either piecewise constant images or piecewise constant images with little noise added and small level of denoising λ. In each of these cases, it is notable that changing the (constant) value of a region can have repercussions on this region and on its immediate neighboring regions only. The propagation of the values change cannot go further.

Another case, more interesting here, is treated in [12]. The authors still assume that the images are defined on a continuous space, and consider an image v that is a linear combination of characteristic functions of convex sets C_i, i.e.,

$$v = \sum_{i=1}^{n} b_i \mathbb{1}_{C_i}, \tag{18}$$

where $(b_i)_{1\leq i\leq n}$ is a sequence of arbitrary real numbers, the sets C_i are assumed to be regular enough and spaced out enough (see [12] for the exact technical assumptions), and each function $\mathbb{1}_{C_i}$ is defined as in (15). Then they prove that the associated denoised image can be written as

$$T(v) = \sum_{i=1}^{n} f_{C_i}(b_i)\mathbb{1}_{C_i},$$

where f_{C_i} is a soft-thresholding function (i.e., an odd function defined by $f_{C_i}(x) = \max(x - \tau_{C_i}, 0)$ for $x \geq 0$) whose threshold τ_{C_i} depends only on C_i. This means that for such images, the convex sets (C_i) evolve independently of each other.

As the operator T is monotone in this continuous framework (see [23] and references therein), we can state a more general property: assume that the image v is null outside convex sets C_i, i.e.,

$$v = \sum_{i=1}^{n} v_i,$$

where each v_i is a bounded image supported by C_i, and the C_i's are regular and spaced out enough as in (18) [12]. The monotony of T implies that

$$T\left(\sum_{i=1}^{n}(\min v_i)\mathbb{1}_{C_i}\right) \leq T(v) \leq T\left(\sum_{i=1}^{n}(\max v_i)\mathbb{1}_{C_i}\right),$$

where both the leftmost and rightmost terms are exactly computable because the arguments of T are in the general form (18). Then $T(v)$ satisfies

$$\sum_{i=1}^{n} f_{C_i}(\min v_i)\mathbb{1}_{C_i} \leq T(v) \leq \sum_{i=1}^{n} f_{C_i}(\max v_i)\mathbb{1}_{C_i},$$

which ensures that $T(v)$ vanishes outside from the sets C_i. Hence the minimization of the ROF energy relative to v holds on the set of images u written as $u = \sum_{i=1}^{n} u_i$ with $\operatorname{supp}(u_i) \subset C_i$.

On such images, the ROF energy can be decomposed into

$$\|u-v\|^2+\lambda TV(u)=\sum_{i=1}^{n}\left(\|u_i-v_i\|^2+\lambda TV(u_i)\right),$$

which is a sum of independent ROF energies, showing that a change in v_i will affect only the component of the energy corresponding to C_i, and that $T(v)$ will then be changed on C_i at most: the components of v evolve independently.

2.3. Locality for natural images: Influence map of a pixel. Here we focus on real-life natural images and show locality properties of TV-denoising. To investigate the locality issue precisely, we shall say that a pixel y is *influenced* by a pixel x if $T(v)(y)$ depends on $v(x)$, and we measure this dependency by

$$C_x(y)=\sup_{\delta\neq 0}\left|\frac{T(v+\delta\mathbb{1}_{\{x\}})(y)-T(v)(y)}{\delta}\right|. \tag{19}$$

This number measures the maximum relative impact on $T(v)(y)$ caused by a distortion of $v(x)$. Several properties of the influence map C_x are gathered in the following theorem.

Theorem 1 (properties of the influence map). *Let Ω be a bounded subset of $\mathbb{Z}^2$, and let $v:\Omega\to\mathbb{R}$. If TV is computed with the ℓ^1-scheme (6) and T is the associated TV-denoising operator, then for any $x\in\Omega$, the influence map C_x defined by (19) satisfies*

(i) $C_x(x)=1$,

(ii) *for every $y\neq x$, $C_x(y)<1$.*

(iii) *for any $\delta\neq 0$,*

$$\sum_{y\in\Omega}\left(\frac{T(v+\delta\mathbb{1}_{\{x\}})(y)-T(v)(y)}{\delta}\right)^2\leq 1.$$

Theorem 1 tells us that a distortion of $v(x)$ cannot be amplified by T, and that it is indeed attenuated for all pixels except x (as we shall see in the experiments, $C_x(y)$ decreases very quickly when y goes away from x). Property (3) does not bring a direct estimate of the total influence of a pixel x but shows that *any given distortion* δ on $v(x)$ causes a total distortion on $T(v)$ smaller than $|\delta|$ (in ℓ^2-norm), since

$$\|T(v+\delta\mathbb{1}_{\{x\}})-T(v)\|_2\leq|\delta|.$$

Proof of Theorem 1. Let $x\in\Omega$. For every $\delta\in\mathbb{R}$, we set $\hat{u}_\delta=T(v+\delta\mathbb{1}_{\{x\}})$, so that

$$C_x(y)=\sup_{\delta\neq 0}|(\hat{u}_\delta(y)-\hat{u}_0(y))/\delta|.$$

We first prove that for every $y\in\Omega$, $C_x(y)\leq 1$. Indeed, if $\delta\neq 0$,

$$v-|\delta|\leq v+\delta\mathbb{1}_{\{x\}}\leq v+|\delta|,$$

and the monotony of T, combined with shift invariance (11), implies that

$$\hat{u}_0-|\delta|\leq\hat{u}_\delta\leq\hat{u}_0+|\delta|.$$

Hence for any pixel $y \in \Omega$,

$$\left| \frac{\hat{u}_\delta(y) - \hat{u}_0(y)}{\delta} \right| \leq 1,$$

and $C_x(y) \leq 1$ by considering the supremum on δ.

Now we come to the very proof of the theorem, beginning with (i), then (iii), and finally (ii).

(1) The Euler equation corresponding to the TV-denoising of image $v + \delta \mathbb{1}_{\{x\}}$ is

$$\hat{u}_\delta - (v + \delta \mathbb{1}_{\{x\}}) + \frac{\lambda}{2} \partial TV(\hat{u}_\delta) \ni 0,$$

where ∂TV denotes the subdifferential of TV, defined by

$$w \in \partial TV(u) \quad \Longleftrightarrow \quad \forall v \in \mathbb{R}^\Omega, \quad TV(v) \geq TV(u) + \langle w, v - u \rangle$$

(see [56] for instance). Let sign be the set-valued function defined by

$$\operatorname{sign}(x) = \begin{cases} \{1\} & \text{if } x > 0, \\ \{-1\} & \text{if } x < 0, \\ [-1, 1] & \text{if } x = 0. \end{cases}$$

The subdifferential of TV, for the ℓ^1-norm of the gradient, is given by

$$\forall u \in \mathbb{R}^\Omega, \ \forall x \in \Omega, \quad \partial TV(u)(x) = \sum_{y \in \Omega, |y-x|_1 = 1} \operatorname{sign}(u(x) - u(y)),$$

and is hence included in $[-4, 4]$. This implies that for each pixel $y \in \Omega$, the gray level $\hat{u}_\delta(y)$ satisfies

$$\hat{u}_\delta(y) \in [v(y) + \delta \mathbb{1}_{\{x\}}(y) - 2\lambda, v(y) + \delta \mathbb{1}_{\{x\}}(y) + 2\lambda].$$

For $y = x$, we get

$$\hat{u}_\delta(x) \in [v(x) + \delta - 2\lambda, v(x) + \delta + 2\lambda],$$

which yields, when δ goes to $+\infty$,

$$\frac{\hat{u}_\delta(x) - \hat{u}_0(x)}{\delta} \geq \frac{v(x) - \hat{u}_0(x) - 2\lambda}{\delta} + 1 \xrightarrow[\delta \to +\infty]{} 1,$$

so that $C_x(x) \geq 1$. Now as $C_x(x) \leq 1$ (as shown above), we get the desired result.

(3) As T is a proximity operator associated to the convex function TV, it is nonexpansive [30, 56]; that is,

$$\forall v_1, v_2 \in \mathbb{R}^\Omega, \quad \|T(v_2) - T(v_1)\|_2 \leq \|v_2 - v_1\|_2. \tag{20}$$

Applying this inequality to $v_1 = v$ and $v_2 = v + \delta \mathbb{1}_{\{x\}}$ yields

$$\sum_{y \in \Omega} (\hat{u}_\delta(y) - \hat{u}_0(y))^2 \leq \delta^2,$$

which proves the desired result.

(2) Again because T is a proximity operator, it is maximal-monotone [56], which means that

$$\forall v_1, v_2 \in \mathbb{R}^\Omega, \quad \langle T(v_2) - T(v_1), v_2 - v_1 \rangle \geq 0,$$

where $\langle f, g \rangle = \sum_{x \in \Omega} f(x)g(x)$ denotes the usual inner product. Let b be an arbitrary real number, and set $v_1 = v$ and $v_2 = v + \delta \mathbb{1}_{\{x\}} + b$. The maximal-monotone property and the shift-invariance of T (11) yield

$$\left\langle \hat{u}_\delta + b - \hat{u}_0, \delta \mathbb{1}_{\{x\}} + b \right\rangle \geq 0.$$

Expanding the inner product leads to

$$\left\langle b\mathbb{1}_\Omega, \delta \mathbb{1}_{\{x\}} + b \right\rangle + \langle \hat{u}_\delta - \hat{u}_0, b\mathbb{1}_\Omega \rangle + \delta(\hat{u}_\delta(x) - \hat{u}_0(x)) \geq 0,$$

and thanks to the average conservation property (10) of T we get

$$|\Omega| b^2 + 2\delta b + \delta(\hat{u}_\delta(x) - \hat{u}_0(x)) \geq 0.$$

This inequality is true for any value of b; hence the discriminant of the quadratic polynomial is nonpositive, which leads, after simplifications, to

$$\forall \delta \neq 0, \quad \left| \frac{\hat{u}_\delta(x) - \hat{u}_0(x)}{\delta} \right| \geq \frac{1}{|\Omega|}.$$

Now using item (3) of the theorem, we can write for any $\delta \neq 0$

$$\sum_{y \neq x} \left(\frac{\hat{u}_\delta(y) - \hat{u}_0(y)}{\delta} \right)^2 \leq 1 - \left(\frac{\hat{u}_\delta(x) - \hat{u}_0(x)}{\delta} \right)^2 \leq 1 - \frac{1}{|\Omega|^2}.$$

This implies that, for any $y \neq x$,

$$C_x(y)^2 \leq 1 - \frac{1}{|\Omega|^2},$$

which ends the proof. ∎

2.4. Experiments. Using numerical experiments, we can investigate some properties of the influence map that are not considered in Theorem 1. To compute C_x numerically on 8-bit images, we use the approximation

$$C_x(y) \simeq \max_{\delta \in \{\pm 2^n, 0 \leq n \leq 9\}} \left| \frac{T(v + \delta \mathbb{1}_{\{x\}})(y) - T(v)(y)}{\delta} \right|,$$

where the TV-denoising operator T is estimated with a very high precision using Chambolle's dual algorithm (the "max" version of [22], mentioned in [23]).

A striking property of the influence map, visible in Figure 1, is the fast decrease of $C_x(y)$ as y moves away from x. For typical values of λ and most pixels x, $C_x(y)$ is below 0.01 as soon as $\|y - x\|$ is larger than 15 pixels. At this point, a natural question arises: is the support of an influence map much smaller than the image domain in general? Following the discussion

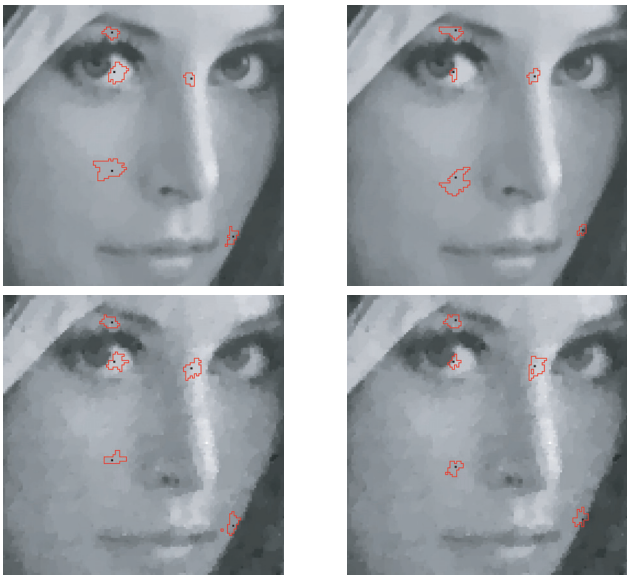


Figure 1. *Influence maps of* 5 *pixels of Lena image after denoising. The two columns correspond to slightly different pixel positions, whereas each row corresponds to different noise and denoising levels (top row:* $\sigma = 10$, $\lambda = 20$*; bottom row:* $\sigma = 25$, $\lambda = 40$*). On each image, the* 5 *considered pixels are marked in black, and the boundary of their significant support* ($\varepsilon = 0.01$) *is marked in red. As we can see, the influence map of each pixel is very concentrated, which suggests that TV-denoising is not far from being a local operator. The precise shape of the influence map, however, seems difficult to predict as it may vary a lot with respect to the point position.*

just before section 2.1, it is reasonable to think that the answer is no for Neumann boundary conditions (used in practice), due to the average-preserving property of TV-denoising in that case. However, in the case of an infinite domain or Dirichlet boundary conditions (section 2.1), there are good reasons to think that, in general, most (it not all) influence maps have a small support. Since the exact support of C_x is difficult to compute in numerical experiments, and in order to neglect the slight global effect resulting from the average preservation induced by Neumann boundary conditions, we propose considering a *significant* support of C_x, defined by

$$\mathrm{supp}_\varepsilon(C_x) = \{y \in \Omega,\, C_x(y) > \varepsilon\},$$

where ε is a small positive threshold ($\varepsilon = 0.01$ in practice). The boundary of this significant support is displayed in Figure 1 (red curves) for some pixels (in black) of the Lena image.

One could wonder whether the support of the influence map could implicitly define a kind of adaptive neighborhood for each pixel. Apart from the fact that the support seems in general larger in flat zones, and partly correlated to the level lines of the denoised images, the level of adaptivity seems to be small in comparison, for example, to the adaptive neighborhoods used in the shape-adaptive discrete cosine function (DCT) (SA-DCT) filter [41, Fig. 2] or in steering kernel regression [61, Fig. 11]. Moreover, the shape of the influence map appears to vary a lot from one pixel to another, even for very near positions, as illustrated by the comparison of the two columns in Figure 1. It is likely that the reciprocal influence map (that

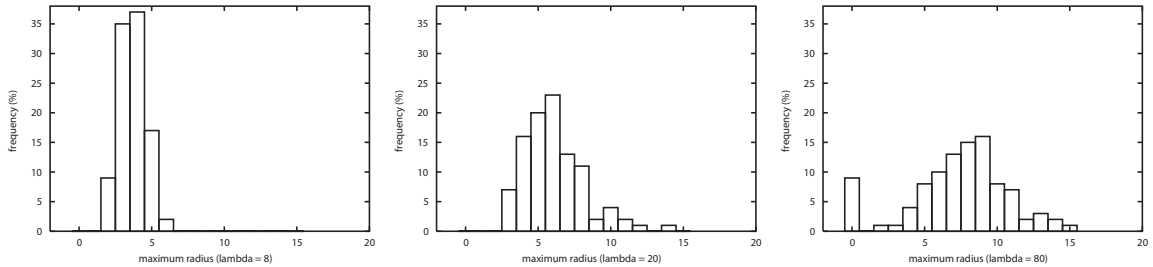


Figure 2. *Given a (randomly chosen) pixel x, we numerically compute the significant support* $\mathrm{supp}_\varepsilon$ *($\varepsilon = 0.01$) of the influence map C_x and measure its radius, that is, the maximum Euclidean distance from x to a point of* $\mathrm{supp}_\varepsilon(C_x)$. *This operation is performed on* 100 *random pixels of a noisy Lena image ($\sigma = 10$) and yields a histogram of the* 100 *observed radii for* 3 *different values of the denoising level (left column: $\lambda = 8$; middle column: $\lambda = 20$; right column: $\lambda = 80$). We can notice that as λ increases, most pixels increase their influence zone (the significant support of C_x gets larger), but the influence never propagates beyond* 15 *pixels. Moreover, for $\lambda = 80$ (right column),* 9% *of the pixels have an influence map with a* 1*-pixel significant support (radius* 0*), which means that these pixels do not significantly influence pixels other than themselves. This phenomenon is likely to be a consequence of the staircasing effect discussed in section* 6.3.

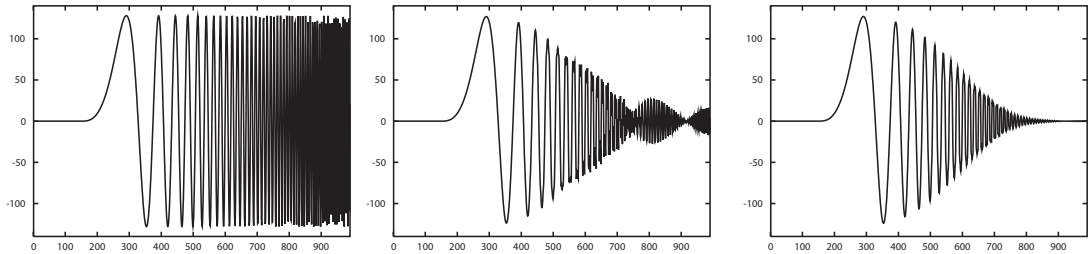
is, the function $x \mapsto C_x(y)$ for a given y, which measures the influence of all pixels x on a given pixel y) would behave in the same way, though we did not test this numerically.

For a more systematic statistical evaluation, we computed the significant support of 100 random pixels of a Lena image and reported in Figure 2 the histograms of their associated radii (maximum distance from x to a point of the significant support $\mathrm{supp}_\varepsilon(C_x)$) for three different values of the regularization parameter λ. As we can see, all significant supports are small, and they tend to grow when λ increases (as for an averaging filter, whose smoothing effect increases with the size of the neighboring window).

In this section, we have given theoretical and numerical evidence that TV-denoising is very near to being a local operator. In particular, we observed that for usual values of the denoising parameter λ, the influence of a given pixel was generally limited to a range of 10–15 pixels. It could be interesting to further study the locality of TV-denoising and, in particular, possibilities of computing exactly the true support of the influence map of a given pixel for appropriate boundary conditions. However, since there probably exists no uniform bound (with respect to pixels and images) on the maximal influence range, it becomes logical at this point, considering the discussion made in the introduction, to reformulate the original variational formulation of TV-denoising in the context of local filters, leading to the next section to what we call *local TV-denoising*.

3. Local TV-denoising.

3.1. Motivation and definition. In this section, we build a purely local filter inspired from TV-denoising: each pixel is processed using the pixels lying in its neighborhood. As we shall see, this construction opens interesting possibilities, such as deriving specific properties of local filters for an operator close to classical TV-denoising (section 4), obtaining a TV-denoising method avoiding the "staircasing" artifact (section 6), and prefiltering the patches of an image efficiently before a patch-based denoising method (section 7).



original signal (chirp) | local TV-denoising with a hard window (13 pixels) | local TV-denoising with a smooth window (same variance)

Figure 3. *Effect of local TV-denoising on the chirp. A chirp signal (left) is denoised with the local TV-filter associated to a neighborhood of size* 13 *(discrete interval), with* $\lambda = 500$ *(middle). Contrary to what could be legitimately expected, frequencies are not attenuated in a monotone way, and low-frequency "waves" (interferences) can be observed on the right part of the signal. This aliasing-like phenomenon, which would appear in a similar way with a boxcar convolution, can be strongly reduced by using an appropriate weight function in the fidelity term (*L^2*-norm) of the variational formulation. This improvement can be observed on the right, where the weighted local TV-denoising introduced in this section has been applied to the chirp signal with a Gaussian weight function of variance* $a^2 = 14$ *(same variance as the hard window) truncated to a large enough window (*21 *pixels).*

Definition 1. *We shall denote as* neighborhood shape *any finite subset* $\mathcal{W}$ *of* $\mathbb{Z}^2$ *that contains* 0.

The most usual neighborhood shapes will be connected sets and in particular rectangles or discrete balls, but it is interesting to notice that the construction we make here does not require any particular geometric assumption on the neighborhood used.

For any pixel $x \in \Omega$, we denote $\mathcal{W}_x = (x + \mathcal{W}) \cap \Omega$ as the neighborhood of x and denote $v(\mathcal{W}_x) \in \mathbb{R}^{\mathcal{W}_x}$ as the image v restricted to this neighborhood. We propose considering the denoising operator $T_\lambda^{\mathcal{W}}$ defined by

$$\forall v \in \mathbb{R}^\Omega,\ \forall x \in \Omega, \quad T_\lambda^{\mathcal{W}}(v)(x) = u_x(x), \tag{21}$$

$$\text{where } \forall x \in \Omega,\ u_x \in \mathbb{R}^{\mathcal{W}_x} \text{ minimizes } \|u_x - v(\mathcal{W}_x)\|^2 + \lambda TV(u_x). \tag{22}$$

Of course here, if $w \in \mathbb{R}^{\mathcal{W}_x}$, $\|w\|$ denotes the restricted ℓ^2-norm $(\sum_{y \in \mathcal{W}_x} w(y)^2)^{\frac{1}{2}}$, and $TV(w)$ is defined from (5) with the boundary conditions arising from the fact that w is defined on $\mathcal{W}_x$. Hence, $T_\lambda^{\mathcal{W}}$ is a local filter, as it amounts to minimizing a local ROF energy on every neighborhood $\mathcal{W}_x$ and keeping the central pixel value $u_x(x)$ only. It is also associated to TV-denoising because letting $\mathcal{W} = \mathbb{Z}^2$ leads to the classical global TV-denoising.

3.2. Window weighting for artifact-free local TV-denoising. Experiments carried out on local TV-denoising with large regularizing parameters λ immediately reveal serious aliasing-like artifacts, that is, the artificial emergence of low frequencies. This phenomenon is particularly visible on signals (which can be considered as images made of a single row to extend previous notation). For instance, in Figure 3, a chirp signal (left) is denoised using local TV-denoising and the result (middle) dramatically suffers from this artifact.

This aliasing-like artifact can be explained by the convergence of local TV-denoising towards linear filtering as the regularity parameter λ tends to infinity. Indeed, as will be seen

in Proposition 2 at the end of this section,

$$T_\lambda^{\mathcal{W}}(v)(x) \xrightarrow[\lambda\to\infty]{} \frac{1}{|\mathcal{W}|}\sum_{y\in\mathcal{W}_x} v(y),$$

so that $T_{\mathcal{W},\lambda}$ is asymptotically equivalent to a discrete convolution with a boxcar function, whose Fourier transform (cardinal sine function) is very oscillating. This explains the aliasing-like artifact for large λ (we use the term "aliasing-like" because it is the interference between the boxcar fundamental frequency and the chirp local frequency that creates a low-frequency envelope).

By analogy to linear filtering, we propose introducing weights on the window $\mathcal{W}$ to attenuate this artifact.

Definition 2 (local weighted TV-denoising). *Let $\omega = (\omega_x) \in \mathbb{R}^{\mathcal{W}}$ be positive weights associated to the neighborhood shape $\mathcal{W}$, and let $\|\cdot\|_\omega$ denote the weighted Euclidean norm defined by*

$$\forall u \in \mathbb{R}^{\mathcal{W}_x}, \qquad \|u\|_\omega^2 = \sum_{y\in\mathcal{W}_x} \omega_{y-x}\, u(y)^2. \tag{23}$$

Then, the local weighted TV-denoiser $T_{\omega,\lambda}$ is an operator on $\mathbb{R}^\Omega$ defined by

$$\begin{aligned} &\forall v \in \mathbb{R}^\Omega,\ \forall x \in \Omega, \qquad T_{\omega,\lambda}(v)(x) = u_x(x), \\ &\text{where } u_x \in \mathbb{R}^{\mathcal{W}_x} \text{ minimizes } E_{\lambda,\omega,x}(u_x) = \|u_x - v(\mathcal{W}_x)\|_\omega^2 + \lambda TV(u_x). \end{aligned} \tag{24}$$

Proposition 2 below gives the asymptotic behavior of the local weighted TV-denoiser when λ goes to $+\infty$.

Proposition 2. *For any positive weight function $\omega \in \mathbb{R}^{\mathcal{W}}$, one has*

$$\forall v \in \mathbb{R}^\Omega,\ \forall x \in \Omega, \qquad T_{\omega,\lambda}(v)(x) \xrightarrow[\lambda\to+\infty]{} \frac{\sum_{y\in\mathcal{W}_x} \omega_{y-x} v(y)}{\sum_{y\in\mathcal{W}_x} \omega_{y-x}}. \tag{25}$$

Proof. A point $x \in \Omega$ being fixed, we denote $\bar{u}$ as the local weighted average around x of an image u, that is,

$$\forall u \in \mathbb{R}^{\mathcal{W}_x}, \qquad \bar{u} = \frac{\sum_{y\in\mathcal{W}_x} \omega_{y-x} u(y)}{\sum_{y\in\mathcal{W}_x} \omega_{y-x}}. \tag{26}$$

If u_λ is the minimizer of $E_{\lambda,\omega,x}$, then 0 is the minimizer of $J : \mathbb{R} \to \mathbb{R}$ defined by

$$J(t) = E_{\lambda,\omega,x}(u_\lambda + t\mathbb{1}) = \|u_\lambda + t\mathbb{1} - v(\mathcal{W}_x)\|_\omega^2 + \lambda TV(u_\lambda),$$

(where $\mathbb{1}$ denotes the constant image equal to 1 everywhere), and writing $J'(0) = 0$ immediately leads to $\bar{u}_\lambda = \bar{v}$, which proves that T preserves average. Now, denoting by C the constant $E_{\lambda,\omega,x}(0)$ independent of λ, we have, for any $\lambda > 0$, $\lambda TV(u_\lambda) \le C$ and $\|u_\lambda - v(\mathcal{W}_x)\|_\omega^2 \le C$. Hence the sequence (u_λ) is uniformly bounded, and the limit u_∞ of any of its subsequences satisfies $TV(u_\infty) = 0$ and thus is a constant. Now, as $T_{\omega,\lambda}$ preserves the average, we have

$\bar{u}_\lambda = \bar{v}$ for every $\lambda > 0$, and passing to the limit (in finite dimension), it follows that $u_\infty = \bar{v}$, which is the only possible limit; this proves that (u_λ) converges to $\bar{v}$. ■

Proposition 2 states that $T_{\omega,\lambda}$ is asymptotically equivalent (when λ goes to $+\infty$) to a linear convolution by the finitely supported kernel $\tilde{\omega} : x \mapsto \omega_{-x}$. Another viewpoint is to let the global contrast of the image go to 0 (which is equivalent to the case $\lambda \to +\infty$ thanks to the change of variable $v \mapsto \alpha v$), and this leads, for a fixed λ, to

$$T_{\omega,\lambda}(\alpha v)(x) \underset{\alpha \to 0}{\sim} \alpha \frac{\sum_{y \in \mathcal{W}_x} \omega_{y-x} v(y)}{\sum_{y \in \mathcal{W}_x} \omega_{y-x}}. \tag{27}$$

Hence, we expect the local TV-denoising to behave like a linear filter in low-contrasted regions or when λ is large (indeed, we shall see later in section 4.4.1 that this linear regime is actually attained for a finite λ).

Now, coming back to the aliasing-like artifacts noticed earlier, it is well known that a linear convolution will avoid aliasing-like artifacts for smooth kernels that have a nonoscillating (unimodal) Fourier transform, and in particular for Gaussian weights defined by

$$\forall x \in \mathcal{W}, \qquad \omega_x = \exp\left(-\frac{|x|^2}{2a^2}\right) \qquad (\text{with } a > 0). \tag{28}$$

Figure 3 (right) shows the result obtained after applying local weighted TV-denoising (with the Gaussian weights given by (28)) on the chirp signal, with an algorithm that will be detailed in section 5. As expected, the aliasing-like artifacts are removed: the signal envelope, which presented artificial oscillations after local TV-denoising with a hard window, has a regular decay when a smooth window is used.

Notice, however, that the aliasing artifacts we mentioned above are probably not so frequent, and avoiding them may not be the most desirable strategy for certain signal or image processing tasks. In particular, the use of a smooth window significantly increases the computational cost of local TV-denoising (see Theorem 4 in section 5.2). It is also worth mentioning that the boxcar filter corresponding to the linear regime of local TV-denoising with a hard window achieves optimal signal denoising in locally constant regions (in terms of signal-to-noise ratio, and among linear convolutions with fixed support kernel).

4. Properties.

In this section, we investigate several properties of the aforementioned local TV-denoising filter, in particular local and global comparison principles and asymptotic behaviors.

4.1. Local comparison.

An important reason to want a local filter instead of a global minimization process is to be able to control the denoising process more locally and more accurately. Among interesting properties is the local comparison principle, which offers a guarantee of local stability. This principle does not always hold for global TV-denoising, but the following proposition states that it does for local TV-denoising (weighted or unweighted).

Proposition 3 (local comparison principle). *Let TV be defined in* (3) *with either an* ℓ^1*- or an* ℓ^2*-norm of the gradient* ((6) *or* (9)). *Then the local filter* $T_{\omega,\lambda}$ *applied on an image* v *satisfies*

$$\forall x \in \Omega, \qquad \min_{y \in \mathcal{W}_x} v(y) \leq T_{\omega,\lambda}\, v(x) \leq \max_{y \in \mathcal{W}_x} v(y).$$

The value of the denoised image hence lies in the local dynamic range of the noisy image. This stability property also echoes the fact that the local TV-denoiser has a more limited denoising level (when $\lambda \to +\infty$, see (25)) than global TV-denoising (for which the most denoised image is a global constant).

Proof. Assume that there exists $x \in \Omega$ such that $T_{\omega,\lambda}(v)(x) \notin [m, M]$, where $m = \min_{y\in\mathcal{W}_x} v(y)$ and $M = \max_{y\in\mathcal{W}_x} v(y)$. Let $u_x \in \mathbb{R}^{\mathcal{W}_x}$ be the patch associated to the denoising of $v(\mathcal{W}_x)$, i.e., minimizing (24). Now let u'_x be the patch with saturated values defined by

$$\forall y \in \mathcal{W}_x, \qquad u'_x(y) = \begin{cases} m & \text{if } u_x(y) \le m, \\ u_x(y) & \text{if } m \le u_x(y) \le M, \\ M & \text{if } M \le u_x(y). \end{cases}$$

Then $\|u'_x - v(\mathcal{W}_x)\|^2_\omega < \|u_x - v(\mathcal{W}_x)\|^2_\omega$ since the gray levels $u_x(y)$ lying away from $[m, M]$ are changed into values which are closer to the noised value $v(y)$. Besides, as

$$\forall y, z \in \mathcal{W}_x, \qquad |u'_x(z) - u'_x(y)| \le |u_x(z) - u_x(y)|,$$

(because $u_x \mapsto u'_x$ is 1-Lipschitz), we have $TV(u'_x) \le TV(u_x)$. This implies that $E_{\lambda,\omega,x}(u'_x) < E_{\lambda,\omega,x}(u_x)$, which contradicts the minimality of $E_{\lambda,\omega,x}(u_x)$. ∎

4.2. Monotony. A global comparison principle that is desirable for general image denoising is the following: if two noisy images v_1 and v_2 satisfy the (pointwise) inequality $v_1 < v_2$, then it would be expected that the related denoised images satisfy a similar large inequality $T(v_1) \le T(v_2)$. This so-called monotony property is an interesting property, since it extends the stability behavior from smooth (say, Lipschitz or more) to nonsmooth images, as a nonsmooth image can always be bounded from below and above by two smooth images.

A convolution filter is monotone as soon as its convolution kernel is nonnegative. Local filters such as the bilateral filter or NL-means, according to our experiments, are generally not monotone (some asymptotic behaviors can be monotone, though). Now for global TV-denoising, it has been shown that the monotony property is true in the continuous framework [4, 5, 20] and in the discrete framework with an ℓ^1-scheme for the gradient norm [23, 35]. However it seems that the monotony does not hold any longer in the discrete framework with an ℓ^2-scheme for the gradient norm (9). This can be seen in the following numerical example made of 3×3-pixel images: if we take

$$v_1 = \begin{array}{|c|c|c|}\hline 42 & 94 & 254 \\ \hline 76 & 178 & 18 \\ \hline 0 & 0 & 0 \\ \hline \end{array} \quad \text{and} \quad v_2 = \begin{array}{|c|c|c|}\hline 43 & 95 & 255 \\ \hline 77 & 179 & 19 \\ \hline 60 & 69 & 105 \\ \hline \end{array},$$

then after a global TV-denoising with an ℓ^2-scheme ($\lambda = 30$), we obtain

$$T(v_1) \approx \begin{array}{|c|c|c|}\hline 60.81 & 98.68 & 224.78 \\ \hline 72.73 & 140.87 & 27.89 \\ \hline 12.08 & 12.08 & 12.08 \\ \hline \end{array} \quad \text{and} \quad T(v_2) \approx \begin{array}{|c|c|c|}\hline 63.29 & 100.49 & 225.65 \\ \hline 83.12 & 138.65 & 60.74 \\ \hline 76.69 & 76.69 & 76.69 \\ \hline \end{array}.$$

Hence we have $v_1 < v_2$ but $T(v_1) \not\le T(v_2)$ (because of the central pixel). This counterexample still holds when the same little images are included in larger constant images set to 0.

Unweighted local TV-denoising directly inherits the possible monotony property of global TV-denoising, as the next proposition states.

Proposition 4 (global comparison principle). *Assume that the global TV-denoising is monotone. Then the local TV-denoising is monotone. In particular, local TV-denoising defined with an ℓ^1-scheme for the gradient norm (6) is monotone.*

Proof. Let $v_1 < v_2$, and let $x \in \Omega$. Let p_1 and p_2 denote the patches that minimize the local ROF energies associated to $v_1(\mathcal{W}_x)$ and $v_2(\mathcal{W}_x)$, respectively. The monotony property of the global TV-denoising operator on $v_1(\mathcal{W}_x) < v_2(\mathcal{W}_x)$ implies that $p_1 \leq p_2$. Hence, denoting $p_i(0)$ as the central gray level of the patch p_i, we get $T_{\mathcal{W},\lambda}v_1(x) = p_1(0) \leq p_2(0) = T_{\mathcal{W},\lambda}v_2(x)$. ∎

Note that it is not difficult to prove that $T_{\mathcal{W},\lambda}$ is continuous, so that $v_1 \leq v_2$ (without a strict inequality) is sufficient to ensure that $T_{\mathcal{W},\lambda}(v_1) \leq T_{\mathcal{W},\lambda}(v_2)$.

The proposed local TV-denoising essentially depends on two parameters: the regularization coefficient λ and the weights (ω_y) which in particular control the locality. As we shall see in the following sections, ω and λ behave in opposite directions. More specifically, when $\lambda \to 0$, the denoising becomes equivalent to the global denoising, and when the neighborhood becomes infinitely small, the denoising becomes equivalent to a linear filtering, as if $\lambda \to +\infty$.

4.3. Asymptotics for $\lambda \to 0$.

4.3.1. Asymptotic equivalence to global TV-denoising when $\lambda \to 0$. In this section we show that local and global TV-denoising have the same asymptotic behavior when $\lambda \to 0$.

Theorem 2. *Let $\mathcal{W}$ be a neighborhood shape and $\omega \in \mathbb{R}^{\mathcal{W}}$ a positive weight function. For any image $v \in \mathbb{R}^{\Omega}$ and any $x \in \Omega$ such that $\mathcal{W}_x$ contains x and its 8 nearest neighbors, and such that TV is continuously differentiable in the neighborhood of $v(\mathcal{W}_x)$, we have*

$$T_{\omega,\lambda}v(x) = v(x) - \frac{\lambda}{2\omega_0}\nabla TV(v)(x) + \underset{\lambda\to 0}{o}(\lambda). \tag{29}$$

Proof. $T_{\omega,\lambda}v(x)$ is the central gray level of the patch $u \in \mathbb{R}^{\mathcal{W}_x}$ which minimizes the energy $E_{\lambda,\omega,x}$ in (24) and whose subdifferential then satisfies

$$\forall y \in \mathcal{W}_x, \qquad 2\omega_{y-x}(u(y) - v(y)) + \lambda \partial TV(u)(y) \ni 0. \tag{30}$$

Let us introduce

$$w_\lambda = \frac{u - v(\mathcal{W}_x)}{\lambda}.$$

Since $E_{\lambda,\omega,x}(v(\mathcal{W}_x)) = \lambda TV(v(\mathcal{W}_x)) \to 0$ when $\lambda \to 0$, we have $\|u - v(\mathcal{W}_x)\|_\omega^2 \to 0$ and thus $\lambda w_\lambda \to 0$ since the weight function ω is positive. Now (30) can be written, for $y = x$,

$$2\omega_0 w_\lambda(x) + \partial TV(v(\mathcal{W}_x) + \lambda w_\lambda)(x) \ni 0.$$

As TV is assumed to be continuously differentiable at the neighborhood of $v(\mathcal{W}_x)$, when λ is small enough, $\partial TV(v(\mathcal{W}_x) + \lambda w_\lambda)(x)$ is equal to $\{\nabla TV(v(\mathcal{W}_x) + \lambda w_\lambda)(x)\}$, and

$$\nabla TV(v(\mathcal{W}_x) + \lambda w_\lambda)(x) \underset{\lambda\to 0}{\longrightarrow} \nabla TV(v(\mathcal{W}_x))(x).$$

Now since the window $\mathcal{W}_x$ contains x and its 8 nearest neighbors (actually 6 neighbors are enough), the restriction to $\mathcal{W}_x$ and the boundary conditions used in (5) have no influence on the terms that depend on x in $TV(v(\mathcal{W}_x))$ and $TV(v)$ (these terms are the same). Consequently, $\nabla TV(v(\mathcal{W}_x))$ and $\nabla TV(v)$ have the same value at point x, and

$$w_\lambda(x) = -\frac{1}{2\omega_0}\nabla TV(v)(x) + \underset{\lambda\to 0}{o}(1)$$

so that $u(x) = v(x) - \frac{\lambda}{2\omega_0}\nabla TV(v)(x) + \underset{\lambda\to 0}{o}(\lambda)$, and the theorem is proved. ■

Remark. In Theorem 2, TV needs to be continuously differentiable in the neighborhood of $v(\mathcal{W}_x)$. As TV is smooth almost everywhere, this is not a strong assumption in practice, since this is true for almost every noisy image v (provided that the noise process admits a density with respect to Lebesgue measure).

In a continuous framework, $-\nabla TV$ corresponds to the curvature operator

$$\operatorname{curv}(u) = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right),$$

and in this case the result of Theorem 2 would be

$$T_{\omega,\lambda}v(x) = v(x) + \frac{\lambda}{2\omega_0}\operatorname{curv}(v)(x) + \underset{\lambda\to 0}{o}(\lambda).$$

This implies that the limiting PDE associated to iterated $T_{\omega,\lambda}$ is

$$\frac{\partial u}{\partial t} = \operatorname{curv} u,$$

which is the PDE associated to global TV-denoising [7, 57], which also corresponds to a degenerate case of the Perona–Malik equation [54].

4.3.2. Application to the normalization of weights. If we want to compare the respective influence of two weight functions ω and ω' for a fixed value of the parameter λ, we need to impose a normalization procedure for the weight functions. In particular, all weight functions $\alpha\omega$ (for $\alpha > 0$) must be normalized into the same weight function, since the local TV-denoising obtained with (ω, λ) and $(\alpha\omega, \lambda/\alpha)$ is exactly the same. Several constraint equations could be used for weight normalization, for example, $\sum_y \omega_y = 1$, $\sum_y \omega_y^2 = 1$, $\frac{1}{|\mathcal{W}|}\sum_y \omega_y = 1$, or $\max_y \omega_y = 1$. A particular normalization is suggested by (29). Indeed, since the asymptotic behavior of the local TV-denoising of a given patch depends only (at first order) on λ/ω_0 and not on the other weight coefficients $(\omega_y)_{y\neq 0}$, it is natural to impose a fixed value of ω_0 as a weight normalization, so that the asymptotic behavior of local TV-denoising depends only on λ (as is the case for global TV-denoising). Hence, in the following all weight functions will be normalized according to the constraint $\omega_0 = 1$, as was done in (28).

4.4. Asymptotics for $\lambda \to +\infty$.

4.4.1. Linear regime reached for finite λ. Here we extend Proposition 2 by showing that the limiting linear regime is reached for a finite λ. This can be seen as the finite counterpart of [24, Lemma 2.3] with an elementary proof.

Proposition 5. *Let $\mathcal{W}$ be a neighborhood shape and $\omega \in \mathbb{R}^{\mathcal{W}}$ a positive weight function. For any image $v \in \mathbb{R}^{\Omega}$ and any $x \in \Omega$, there exists a critical value denoted by $\lambda_c(x)$ such that*

$$\forall \lambda \geq \lambda_c(x), \qquad T_{\omega,\lambda}(v)(x) = \frac{\sum_{y \in \mathcal{W}_x} \omega_{y-x} v(y)}{\sum_{y \in \mathcal{W}_x} \omega_{y-x}}.$$

Proof. Let $\bar{v} = \frac{\sum_{y \in \mathcal{W}_x} \omega_{y-x} v(y)}{\sum_{y \in \mathcal{W}_x} \omega_{y-x}}$. We prove that the constant image $\bar{v}$ achieves the minimum of $E_{\lambda,\omega,x}$ (defined in (24)) by considering the behavior of $E_{\lambda,\omega,x}$ at the neighborhood of $\bar{v}$. Let $u \in \mathbb{R}^{\mathcal{W}_x}$; it can be decomposed into $u = \delta u + \bar{v} + \alpha$, where δu is an image with zero weighted mean and α is a scalar.

First, denoting $\langle \cdot, \cdot \rangle_\omega$ as the inner product associated to the Hilbert norm $\| \cdot \|_\omega$, and $\mathbb{1}$ as the constant image equal to 1 everywhere on $\mathcal{N}_x$, notice that

$$E_{\lambda,\omega,x}(\delta u + \bar{v} + \alpha) - E_{\lambda,\omega,x}(\delta u + \bar{v}) = \|\alpha \mathbb{1}\|_\omega^2 + 2 \langle \alpha \mathbb{1}, \delta u + \bar{v} - v \rangle_\omega = \|\alpha \mathbb{1}\|_\omega^2 \tag{31}$$

since $\delta u + \bar{v} - v$ has zero weighted mean. Second,

$$E_{\lambda,\omega,x}(\delta u + \bar{v}) - E_{\lambda,\omega,x}(\bar{v}\mathbb{1}) = \|\delta u\|_\omega^2 + 2 \langle \bar{v} - v, \delta u \rangle_\omega + \lambda TV(\delta u), \tag{32}$$

and setting $m(w) = \frac{1}{|\mathcal{W}|} \sum_{x \in \mathcal{W}} w(x)$ yields

$$2 \langle \bar{v} - v, \delta u \rangle_\omega = 2 \langle \bar{v} - v, \delta u - m(\delta u) \rangle_\omega \geq -2 \|\omega(\cdot - x)(v - \bar{v})\|_\infty \|\delta u - m(\delta u)\|_1. \tag{33}$$

Now, as all the norms are equivalent in finite dimension, there exists $C > 0$ such that

$$\forall w \in \mathbb{R}^{\mathcal{W}_x}, \quad |m(w)| + TV(w) \geq C \|w\|_1. \tag{34}$$

Applying this inequality to $w = \delta u - m(\delta u)$, we get $TV(\delta u) \geq C \|\delta u - m(\delta u)\|_1$, so that with (32) and (33) we obtain

$$E_{\lambda,\omega,x}(\bar{v} + \delta u) - E_{\lambda,\omega,x}(\bar{v}\mathbb{1}) \geq \|\delta u\|_\omega^2 + (C\lambda - 2\|\omega(\cdot - x)(v - \bar{v})\|_\infty) \|\delta u - m(\delta u)\|_1. \tag{35}$$

Last, adding (31) to (35) yields

$$E_{\lambda,\omega,x}(u) - E_{\lambda,\omega,x}(\bar{v}\mathbb{1}) \geq \|\alpha \mathbb{1}\|_\omega^2 + \|\delta u\|_\omega^2 + (C\lambda - 2\|\omega(\cdot - x)(v - \bar{v})\|_\infty) \|\delta u - m(\delta u)\|_1,$$

and the right-hand term is nonnegative for any $u \in \mathbb{R}^{\mathcal{W}}$ as soon as $\lambda \geq 2\|\omega(\cdot - x)(v - \bar{v})\|_\infty / C$. It means that for large enough values of λ, $\bar{v}\mathbb{1}$ reaches the minimum of $E_{\lambda,\omega,x}$. ■

4.4.2. Asymptotic behavior for small neighborhoods. Here we show that in a continuous setting, letting the size of the neighborhood $\mathcal{W}$ go to 0 (while keeping λ constant) is equivalent to letting $\lambda \to +\infty$, hence asymptotically leading to linear filtering.

We first define local TV-denoising in a continuous setting. Let Ω denote an open subset of $\mathbb{R}^2$, and let $\mathcal{W}$ be a bounded convex open subset of $\mathbb{R}^2$ containing 0, weighted by a positive function $\omega \in L^\infty(\mathcal{W})$. If $x \in \Omega$, the set $\mathcal{W}_x = (x + \mathcal{W}) \cap \Omega$ is a neighborhood of x. The total variation of a patch $u \in L^2(\mathcal{W}_x)$ is defined by duality by

$$TV(u) = \sup \left\{ \int_{\mathcal{W}_x} u \operatorname{div} p,\ p \in \mathcal{C}_c^\infty(\mathcal{W}_x, \mathbb{R}^2),\ \|p\|_\infty \leq 1 \right\} \tag{36}$$

(see [8]), and its weighted norm $\|u\|_\omega$ by

$$\|u\|_\omega^2 = \int_{\mathcal{W}_x} \omega(y - x)\, u(y)^2\, dy. \tag{37}$$

Given $\mathcal{W}$ and $\lambda > 0$, we consider the unique function u belonging to

$$BV(\mathcal{W}_x) = \{u \in L^2(\mathcal{W}_x),\ TV(u) < +\infty\}$$

that minimizes

$$E_{\lambda,\omega,x}(u) = \|u - v(\mathcal{W}_x)\|_\omega^2 + \lambda TV(u), \tag{38}$$

where $v(\mathcal{W}_x)$ denotes the restriction of v on the subdomain $\mathcal{W}_x$. Then, the local TV-denoising operator $T_{\mathcal{W},\lambda}$ at point x is defined by

$$T_{\mathcal{W},\lambda} v(x) = \lim_{r \to 0^+} \frac{1}{|B(x,r)|} \int_{B(x,r)} u \tag{39}$$

when the limit exists ($B(x,r)$ is the Euclidean open ball with center x and radius r, and $|B(x,r)|$ is its Lebesgue measure).

Note that in (39), we need to consider a limit (the mean value of u at point x) because the value $u(x)$ has no meaning by itself (u is defined up to a Lebesgue-negligible function). Since $u \in BV(\mathcal{W}_x)$, this limit exists almost surely (that is, for almost any point of $\mathcal{W}_x$), which does not prove that $T_{\mathcal{W},\lambda}$ is defined almost everywhere though (even if we believe that it is the case). In the following, the existence of $T_{\mathcal{W},\lambda} v(x)$ will always be ensured in the asymptotic frameworks we consider.

Now, in order to make the neighborhood $\mathcal{W}$ shrink to a singleton, we consider a dilation parameter $h > 0$ and the dilated set $h\mathcal{W}$, associated to the weight function $\omega(\cdot/h)$, in agreement with the normalization suggested in section 4.3.2. The following theorem describes the asymptotic behavior of $T_{h\mathcal{W},\lambda}$ when h goes to 0.

Theorem 3. *Let Ω be an open subset of $\mathbb{R}^2$ and $v \in \mathcal{C}^3(\Omega)$. Let $\mathcal{W}$ be a bounded convex open neighborhood of 0 in $\mathbb{R}^2$ and let $\omega \in L^\infty(\mathcal{W})$ satisfy*

$$\omega > 0, \qquad \int_{\mathcal{W}} \omega(x)\, x\, dx = 0, \qquad \text{and}$$

$$\mathrm{Cov}_{\mathcal{W}}(\omega) := \frac{1}{\int_{\mathcal{W}} \omega} \begin{pmatrix} \int_{\mathcal{W}} x_1^2\, \omega(x)\, dx & \int_{\mathcal{W}} x_1 x_2\, \omega(x)\, dx \\ \int_{\mathcal{W}} x_1 x_2\, \omega(x)\, dx & \int_{\mathcal{W}} x_2^2\, \omega(x)\, dx \end{pmatrix} = \sigma^2 Id$$

(with the usual convention $x = (x_1, x_2)$). For each $x \in \Omega$, when h is small enough, the denoising operator $T_{h\mathcal{W},\lambda}$ considered by (39) is well defined and satisfies

$$T_{h\mathcal{W},\lambda} v(x) = v(x) + \frac{h^2\sigma^2}{2} \Delta v(x) + \underset{h\to 0}{O}(h^3), \tag{40}$$

where $\Delta v = \frac{\partial^2 v}{\partial x_1^2} + \frac{\partial^2 v}{\partial x_2^2}$ denotes the Laplacian of v.

It is quite surprising that a restoration method based on TV, which assigns a finite cost to contrasted edges but favors piecewise constant structures more than smooth structures (staircasing effect), could be associated to an isotropic diffusion through the Laplacian operator. Indeed, local TV-denoising is equivalent to global TV-denoising for large enough neighborhoods (see section 6.1), but Theorem 3 points out that it is equivalent to Gaussian filtering for very small neighborhoods. This can be linked to the linear behavior of local TV-denoising in low-contrasted regions (27). Local TV-denoising with middle-sized neighborhoods hence reaches a compromise between global TV-denoising and Gaussian filtering.

Lemma 1. *Let $v \in L^2(\Omega)$ and $x \in \Omega$. Assume that $\mathcal{W}$ is an open subset of $\mathbb{R}^2$ such that $\mathcal{W}_x$ is bounded and convex, and consider a positive weight function $\omega \in L^\infty(\mathcal{W})$. Let $\bar{v}_x = \frac{\int_{\mathcal{W}_x} \omega(y-x) v(y)\, dy}{\int_{\mathcal{W}_x} \omega(y-x)\, dy}$. If*

$$\lambda \geq \mathrm{diam}(\mathcal{W}_x) \cdot \|\omega(\cdot - x)(v(\mathcal{W}_x) - \bar{v}_x)\|_{L^\infty(\mathcal{W}_x)} \tag{41}$$

(where $\mathrm{diam}(\mathcal{W}_x)$ denotes the diameter of $\mathcal{W}_x$), then the constant image $\bar{v}_x$ minimizes $E_{\lambda,\omega,x}$.

Remark. The existence of such a bound on λ (41) above which the constant image minimizes $E_{\lambda,\omega,x}$ is proved in [24, Lemma 2.3] in a slightly different framework (the authors consider the eventuality of a blurring operator, but $\omega = 1$). Here we derive an explicit upper bound and propose a proof directly inspired from the discrete framework.

Proof of Lemma 1. The proof is similar to that of the discrete framework in Proposition 5. Only (34) has to be justified, with $C = 2/\mathrm{diam}(\mathcal{W}_x)$. Actually, in a continuous framework, the Poincaré inequality [8] states that for some constant $\gamma > 0$, one has

$$\forall u \in L^1(\mathcal{W}_x), \quad \|u - m(u)\|_{L^1(\mathcal{W}_x)} \leq \gamma\, TV(u),$$

where $m(u) = \frac{1}{|\mathcal{W}|} \int_{\mathcal{W}} u(y)\, dy$. In [1], it is shown that if $\mathcal{W}_x$ is bounded and convex, the previous inequality holds with $\gamma = \mathrm{diam}(\mathcal{W}_x)/2 = 1/C$, which completes the proof. ∎

Proof of Theorem 3. For any $h > 0$ and $x \in \Omega$, let $\mathcal{W}_{h,x} = (x + h\mathcal{W}) \cap \Omega$ and

$$\bar{v}_{h,x} = \frac{\int_{\mathcal{W}} \omega(y)\, v(x + ht)\, dt}{\int_{\mathcal{W}} \omega(t)\, dt}.$$

Consider $h_0 > 0$ such that for any $h < h_0$, $\mathcal{W}_{h,x} \subset \Omega$ (and hence is convex). By Lemma 1,

$$\lambda^c_{h,x} = \operatorname{diam}(\mathcal{W}_{h,x}) \cdot \|\omega((\cdot - x)/h)(v(\mathcal{W}_{h,x}) - \bar{v}_{h,x})\|_{L^\infty(\mathcal{W}_{h,x})}$$

is a critical value of λ, that is, a value above which the denoised version of $v(\mathcal{W}_{h,x})$ is constant, equal to $\bar{v}_{h,x}$. Now we have

$$\lambda^c_{h,x} \leq h \operatorname{diam}(\mathcal{W}) \cdot 2\|\omega\|_{L^\infty(\mathcal{W})}\|v(\mathcal{W}_x)\|_{L^\infty(\mathcal{W}_{h_0,x})} \xrightarrow[h\to 0]{} 0;$$

hence there exists $h_1 \in (0, h_0]$ such that $\lambda \geq \lambda^c_{h,x}$ as soon as $h < h_1$. Thus, for any $h < h_1$, the denoising is linear at point x, and $T_{h\mathcal{W},\lambda}v(x) = \bar{v}_{h,x}$.

We end the proof by deriving an asymptotic development of $\bar{v}_{h,x}$ when $h \to 0$. As v is $\mathcal{C}^3$, we can write for any $h \in (0, h_1)$,

$$\forall t \in \mathcal{W}, \qquad -Mh^3 \leq v(x+ht) - v(x) - h\nabla v(x) \cdot t - \frac{h^2}{2}D^2v(x)(t,t) \leq Mh^3, \tag{42}$$

where $M = \frac{1}{6}\sup_{h_1\mathcal{W}} \|D^3v\| \cdot \sup_{h_1\mathcal{W}} |t|^3$ (notice that $M < +\infty$ because D^3v is continuous on the closed set $h_1\overline{\mathcal{W}}$). Multiplying all terms of (42) by $\omega(t)$ and integrating on $\mathcal{W}$, we then obtain

$$\left|(\bar{v}_{h,x} - v(x)) \int_{\mathcal{W}} \omega - h\nabla v(x) \cdot \int_{\mathcal{W}} \omega(t)t\,dt - \frac{h^2}{2}\int_{\mathcal{W}} \omega(t)D^2v(x)(t,t)\,dt\right| \leq Mh^3 \int_{\mathcal{W}} \omega.$$

Using the fact that $\operatorname{Cov}_{\mathcal{W}}(\omega) = \sigma^2 Id$ and $\int_{\mathcal{W}} \omega(t)t\,dt = 0$, we get

$$\int_{\mathcal{W}} \omega(t)D^2v(x)(t,t)\,dt = \sigma^2\Delta v(x)\int_{\mathcal{W}} \omega$$

and finally

$$\bar{v}_{h,x} = v(x) + \frac{h^2\sigma^2}{2}\Delta v(x) + \underset{h\to 0}{O}(h^3)$$

with $T_{h\mathcal{W},\lambda}v(x) = \bar{v}_{h,x}$ for h small enough, as noticed above. ∎

5. Algorithm. In this section, we propose a dual algorithm based on Chambolle's work [22] that achieves the minimization of the weighted energy

$$E_{\lambda,\omega,v}(u) = \sum_{x\in\mathcal{W}} \omega_x(u(x) - v(x))^2 + \lambda TV(u), \tag{43}$$

where $\mathcal{W}$ is a finite arbitrary domain of $\mathbb{Z}^2$, u, v are in $\mathbb{R}^{\mathcal{W}}$, and $(\omega_x) \in \mathbb{R}^{\mathcal{W}}$ are positive weights.

5.1. Characterization of the minimizer. Let $\mathcal{W}$ be an arbitrary subset of $\mathbb{Z}^2$ (not necessarily a rectangular domain). We assume as in [22] that the TV is discretized according to (5), using the ℓ^2-norm (9). In particular, if $y = (y^1, y^2) \in \mathbb{R}^2$, $|y|$ will denote its modulus, i.e.,

$|y| = \sqrt{(y^1)^2 + (y^2)^2}$. In order to introduce the discrete divergence operator, let $\langle \cdot, \cdot \rangle$ denote the usual inner product on $\mathbb{R}^{\mathcal{W}}$, and $\langle \cdot, \cdot \rangle_Y$ the inner product on $(\mathbb{R}^2)^{\mathcal{W}}$ defined by

$$\forall p = (p^1, p^2) \in (\mathbb{R}^2)^{\mathcal{W}}, \quad \forall q = (q^1, q^2) \in (\mathbb{R}^2)^{\mathcal{W}}, \quad \langle p, q \rangle_Y = \sum_{x \in \mathcal{W}} (p_x^1 q_x^1 + p_x^2 q_x^2),$$

the associated Euclidean norms being written as $\| \cdot \|$ and $\| \cdot \|_Y$, respectively. Let $\delta_{i,j}^{\mathcal{W}}$ denote the real 1 if $(i,j) \in \mathcal{W}$ and 0 otherwise. If $p \in (\mathbb{R}^2)^{\mathcal{W}}$, the discrete divergence of p is an image $\operatorname{div} p$ defined by

$$\forall (i,j) \in \mathcal{W}, \quad (\operatorname{div} p)(i,j) = p_{i,j}^1 \delta_{i+1,j}^{\mathcal{W}} - p_{i-1,j}^1 \delta_{i-1,j}^{\mathcal{W}} + p_{i,j}^2 \delta_{i,j+1}^{\mathcal{W}} - p_{i,j-1}^2 \delta_{i,j-1}^{\mathcal{W}} \tag{44}$$

(note that this formulation holds regardless of convention used to define p_x when $x \notin \mathcal{W}$). This divergence operator div is dual to the gradient in the sense that

$$\forall u \in \mathbb{R}^{\mathcal{W}}, \ \forall p \in (\mathbb{R}^2)^{\mathcal{W}}, \ \langle \operatorname{div} p, u \rangle = -\langle p, \nabla u \rangle_Y.$$

We also write $\|p\|_\infty = \max_{x \in \mathcal{W}} |p_x|$ and consider the (invertible) diagonal operator D defined by

$$\forall u \in \mathbb{R}^{\mathcal{W}}, \ \forall x \in \mathcal{W}, \qquad (Du)(x) = \omega_x u(x), \tag{45}$$

so that for any $u \in \mathbb{R}^{\mathcal{W}}$ we have $\|u\|_\omega = \sqrt{\sum_{x \in \mathcal{W}} \omega_x u(x)^2} = \|D^{1/2} u\|$. The next proposition characterizes the minimizer of (24) as the projection on a convex set.

Proposition 6. *The minimizer of* (43) *can be written as*

$$T_\omega v = v - \pi_{\frac{\lambda}{2}K}(v),$$

where

$$K = \left\{ D^{-1}(\operatorname{div} p), \ \ p \in (\mathbb{R}^2)^{\mathcal{W}}, \ \|p\|_\infty \leq 1 \right\} \tag{46}$$

is closed and convex, and $\pi_{\frac{\lambda}{2}K}$ *denotes the projection operator on* $\frac{\lambda}{2}K = \{\frac{\lambda}{2}k, \ k \in K\}$. *Furthermore, if* $p \in (\mathbb{R}^2)^{\mathcal{W}}$ *is such that* $\pi_{\frac{\lambda}{2}K}(v) = \frac{\lambda}{2} D^{-1}(\operatorname{div} p)$, *then* p *is characterized by*

$$\forall x \in \mathcal{W}, \qquad \left| \left(\nabla \left(\frac{\lambda}{2} D^{-1}(\operatorname{div} p) - v \right) \right)_x \right| \ p_x = \left(\nabla \left(\frac{\lambda}{2} D^{-1}(\operatorname{div} p) - v \right) \right)_x . \tag{47}$$

For the sake of completeness, a proof derived from [22] is detailed in Appendix A. Again following [22], we can now derive an iterative scheme for the numerical minimization of (43) by considering a semi-implicit gradient descent with step $\tau > 0$, given by

$$\forall x \in \mathcal{W}, \quad p_x^{n+1} = p_x^n + \tau \left[\nabla \left(D^{-1} \operatorname{div} p^n - \frac{v}{\lambda/2} \right) - \left| \nabla \left(D^{-1} \operatorname{div} p^n - \frac{v}{\lambda/2} \right) \right| p^{n+1} \right]_x , \tag{48}$$

which leads to

$$\forall x \in \mathcal{W}, \qquad p_x^{n+1} = \frac{p_x^n + \tau (\nabla (D^{-1} \operatorname{div} p^n - \frac{v}{\lambda/2}))_x}{1 + \tau |\nabla (D^{-1} \operatorname{div} p^n - \frac{v}{\lambda/2})_x|}. \tag{49}$$

5.2. Convergence.

5.2.1. Main result. When $\mathcal{W}$ is a rectangle and $\omega_x = 1$ for all $x \in \mathcal{W}$ (i.e., $D = Id$), the iterative scheme (49) converges, provided that $\tau \leq 1/8$ [22, Theorem 3.1]. In the case of an arbitrary domain $\mathcal{W}$ and arbitrary positive weights $(\omega_x)_{x\in\mathcal{W}}$, convergence still holds with an adequate bound on the time step given by the following theorem.

Theorem 4. *Let* $(\omega_x)_{x\in\mathcal{W}}$ *be* positive *weights on* $\mathcal{W}$*, and let*

$$\tau_{\max} = \frac{1}{4\max\left\{\max^*_{i,j}\left(\frac{1}{\omega_{i,j}}+\frac{1}{\omega_{i+1,j}}\right), \max^*_{i,j}\left(\frac{1}{\omega_{i,j}}+\frac{1}{\omega_{i,j+1}}\right)\right\}}, \tag{50}$$

with the $*$ *symbol in* $\max^*_{i,j} f(i,j)$ *meaning that only the indices* (i,j) *for which* $f(i,j)$ *is defined are considered. Let* D *be the linear diagonal invertible operator defined by* (45)*,* $v \in \mathbb{R}^{\mathcal{W}}$*, and* $\lambda > 0$*. For any* $\tau \in (0, \tau_{\max}]$*, if* (p^n) *is arbitrarily initialized (with* $\|p^0\|_\infty \leq 1$*) and recursively defined by* (49)*, then* $v - \frac{\lambda}{2}D^{-1}(\operatorname{div} p^n)$ *converges to the minimizer of* $E_{\lambda,\omega,v}$ (43)*.*

This result is a small refinement of a particular case of [3, Theorem 1], itself derived from [22, Theorem 3.1]. The improvements we bring concern the convergence bound $\tau_{\max}$ and the domain (not constrained to be a rectangle). We give a self-contained proof in Appendix B for completeness.

Remark 1 (maximum upper bound). If the maximum weight is set to 1 (which is relevant with the normalization we propose in section 4.3.2 when $\omega_0 = 1$), then $\min_x \omega_x/8 \leq \tau_{\max} \leq 1/8$. Note that the equality $\tau_{\max} = 1/8$ (which is the bound given in [22]) is reached only for uniform weights ($\omega \equiv 1$).

Remark 2 (practical convergence). In numerical experiments where $\omega \equiv 1$, the effective maximal step size allowing convergence is $1/4$ [22], that is, twice the limit $\tau_{\max}$ predicted by the theory (Aujol [9] gives an explanation for this, derived from the Bermúdez–Moreno algorithm). For Gaussian weights (28), we observed that the effective maximum step size $\tau^{\text{eff}}_{\max}$ satisfied

$$\tau^{\text{eff}}_{\max} \in [2\tau_{\max}, 4\tau_{\max}]. \tag{51}$$

The value $\tau^{\text{eff}}_{\max} \simeq 4\tau_{\max}$ is found when $\mathcal{W}$ is much larger than the kernel's bandwidth, so that the Gaussian kernel is hardly truncated by the boundary of $\mathcal{W}$. Conversely, when $\mathcal{W}$ is much smaller than the kernel bandwidth, the weights are virtually uniform on $\mathcal{W}$, and $\tau^{\text{eff}}_{\max} \simeq 2\tau_{\max}$. In the latter case, however, the aliasing-like effect is to be expected (as shown in section 3.2), so that a trade-off has to be found between the complete removal of the artifact and the speed of the algorithm. Note that for the projection variant of (49),

$$p^{n+1}_x = \frac{p^n_x + \tau(\nabla(D^{-1}\operatorname{div} p^n - \frac{v}{\lambda/2}))_x}{\max\left(1, |p^n_x + \tau(\nabla(D^{-1}\operatorname{div} p^n - \frac{v}{\lambda/2}))_x|\right)} \tag{52}$$

which generalizes a scheme proposed in [23] to the weighted case, it is not difficult to prove that convergence holds as soon as $\tau < \min_x \omega_x/4$, which is twice the bound given in [3] (see [9] for a proof derived from the Bermúdez–Moreno algorithm in the nonweighted case).

5.2.2. Pointwise convergence control. Adapting Chambolle's criterion [23] to our case, we are able to explicitly control the weighted L^2-distance between the optimum denoised patch and the patch coming from the iterations of the algorithm. Indeed, let $u^n = v - \frac{\lambda}{2} D^{-1} \operatorname{div} p^n$ be the image obtained after the nth iteration. Let also $\bar{p}$ denote the limit of p^n, and $\bar{u} = v - \frac{\lambda}{2} D^{-1} \operatorname{div} \bar{p}$. Then,

$$\begin{aligned} \|u^n - \bar{u}\|_\omega^2 &= \left\langle \frac{\lambda}{2} D^{-1}(\operatorname{div} \bar{p} - \operatorname{div} p^n), D(u^n - \bar{u}) \right\rangle \\ &= \frac{\lambda}{2} \langle \operatorname{div} \bar{p} - \operatorname{div} p^n, u^n - \bar{u} \rangle \\ &= \frac{\lambda}{2} \langle p^n - \bar{p}, \nabla u^n - \nabla \bar{u} \rangle_Y . \end{aligned}$$

Now $|\langle p, \nabla u \rangle_Y| \leq TV(u)$ for every $u \in \mathbb{R}^{\mathcal{W}}$ and for every $p \in (\mathbb{R}^{\mathcal{W}})^2$ satisfying $\|p\|_\infty \leq 1$. But as $\langle \bar{p}, \nabla \bar{u} \rangle_Y = -TV(\bar{u})$ (consider (47) with $\frac{\lambda}{2} D^{-1} \operatorname{div} \bar{p} = v - \bar{u}$), we get

$$\|u^n - \bar{u}\|_\omega^2 \leq \frac{\lambda}{2} \left(TV(u^n) + \langle p^n, \nabla u^n \rangle_Y \right),$$

which is a computable local convergence bound (the right-hand term going to 0 as $n \to +\infty$) that entails the pointwise convergence criterion

$$\forall x \in \mathcal{W}, \quad |u^n - \bar{u}|^2(x) \leq \frac{\lambda}{2} \left(TV(u^n) + \langle p^n, \nabla u^n \rangle_Y \right) \tag{53}$$

since $\omega_0 = 1$.

5.2.3. Usefulness of an arbitrary neighborhood shape. The first main benefit of an arbitrary neighborhood shape is to speed up the algorithm. Indeed, the pixels x corresponding to negligible values of the weight ω_x can be removed from the neighborhood shape without much change in the associated local TV-filter, and this results in an arbitrary neighborhood shape with an increased value of $\tau_{\max}$ (which speeds up the algorithm).

A second important benefit is to be able to deal with adaptive neighborhoods. A typical shape to be used for $\mathcal{W}$ is a discrete ball with a fixed radius, so that the algorithm becomes more isotropic. But the neighborhoods can also be designed adaptively with respect to the image: not only can the radius be adapted to the local scale of the image, but the shape could also be distorted along the geometry of the image as in [43]. We shall focus on different applications in the remaining part of this paper, but these ideas could be interesting directions to explore.

6. Experiments with local TV-denoising. In this section, the local TV-filter considered in Definition 2 is directly used as a denoising filter. All denoising experiments that follow are computed with the iterative algorithm described in section 5. Basically, this algorithm is not very fast because it requires the minimization of a local energy for each pixel. Hopefully, we can take advantage of the photometric similarity between successive image patches (according to the lexicographic order on pixels) to speed up the minimization process. Indeed, as TV-denoising is nonexpansive (20), the spatial coherence of the image implies that the solution

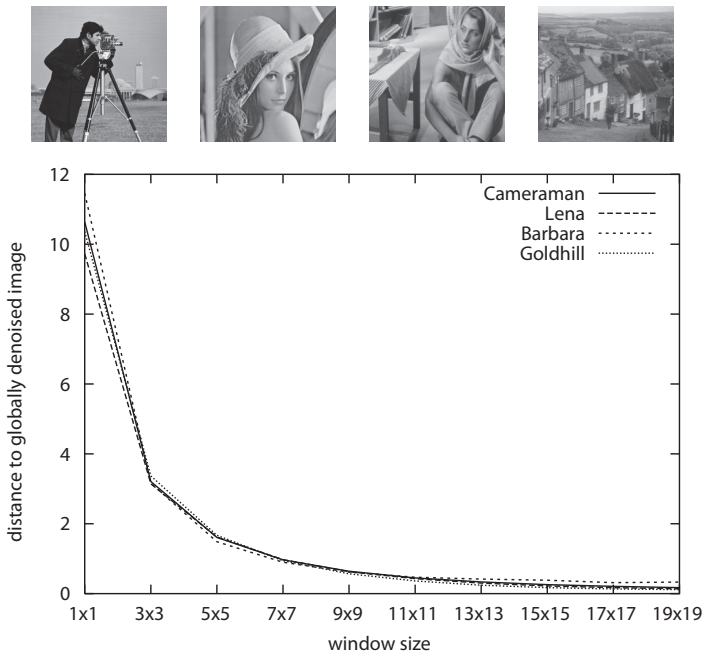


Figure 4. *From local to global denoising. For each of the 4 classical images Cameraman, Lena, Barbara, and Goldhill, degraded by an additive white Gaussian noise ($\sigma = 10$), the graph shows the quick decrease of the L^2-distance between the globally denoised image and the locally denoised one using increasing window sizes (hard window and constant $\lambda = 20$). Despite the fact that the image contents are very different (miscellaneous textures and different scales), the convergence curves above seem to be quite independent of the image.*

associated to a given pixel is necessarily close to the solution associated to its four adjacent pixels. Thus, by initializing the algorithm for each pixel with the solution found for the previous pixel, we start the iterative process with a much better initial guess than a classical initialization (e.g., a constant image), and the convergence is attained much more quickly.

Concerning the processing of image borders, we adopted a classical solution that consists of extending the image domain with a symmetry convention. This permits us to maintain, for each pixel of the original image domain, the same neighborhood shape. Another solution would have been to intersect the theoretical neighborhood of each pixel with the image domain, but such a process is likely to artificially favor a linear denoising near the image borders due to the decrease of the neighborhood area (see section 4.4.2).

6.1. Locality of global TV. Now that a local version of TV-denoising has been proposed, we are able to confirm by numerical experiments that global TV-denoising is mainly ruled by local interactions (see section 2). In Figure 4, several images denoised by local TV are compared using the L^2-norm to those denoised by global TV, for all sizes of square neighborhoods $\mathcal{W}$ smaller than 19×19 (and constant weight functions in all neighborhoods). As predicted in section 2, the locally denoised images quickly converge to the globally denoised ones when the neighborhood size increases: in practice an 11×11 patch is sufficient to capture most of the interactions, which means that long-range interactions caused by global TV-denoising are globally negligible in these ordinary images. It is also interesting to notice that the curves

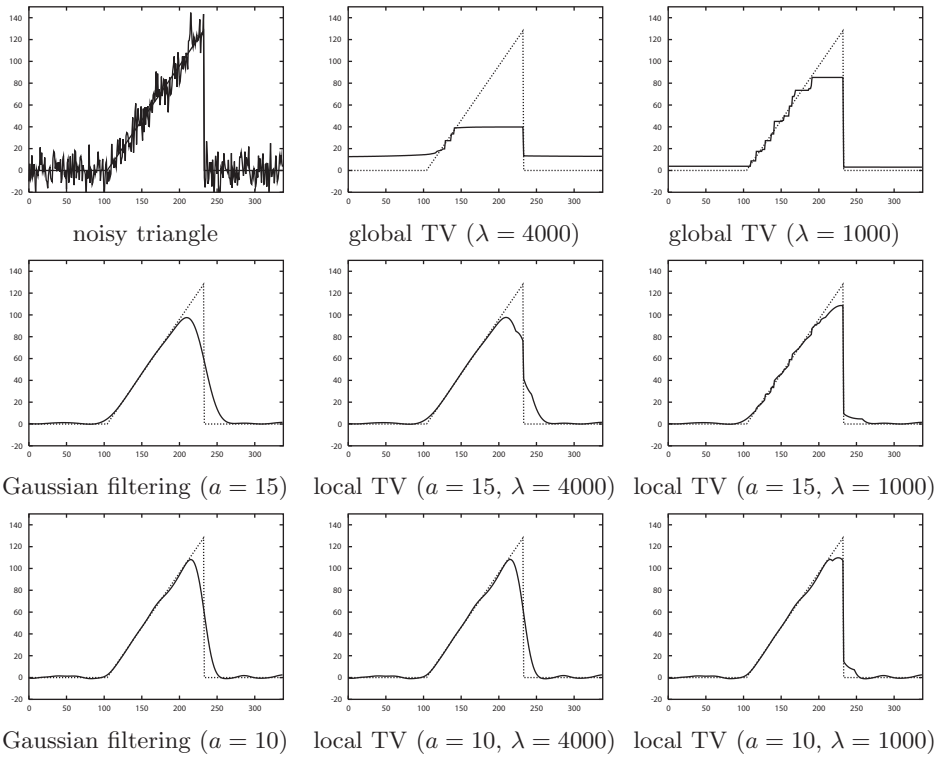


Figure 5. *A triangle signal corrupted by an additive white Gaussian noise with $\sigma = 10$ (top left) is denoised by three different algorithms, which all correspond to particular cases of local TV-denoising. The global TV method (top row), which corresponds to $a = +\infty$, preserves the original "discontinuity" (that is, the presence of an intensity gap) but largely destroys the maximum intensity zone and creates staircasing artifacts. The Gaussian filtering (first column), which corresponds to $\lambda = +\infty$, avoids staircasing and extrema killing but introduces a lot of blur that completely loses the original intensity gap. An interesting compromise between these two methods can be obtained by using local TV-denoising with finite values of a and λ, as can be seen in the four bottom right images (2×2 square) and in particular the two rightmost images ($\lambda = 1000$).*

in Figure 4 more or less coincide. This relative stability among images is likely to make the choice of the window size easier.

6.2. Local versus global denoising, bias-variance trade-off. Consider the local TV-filter associated to a Gaussian weight function with standard deviation a (written ω^a), as in (28). When the locality parameter a is small, local TV-denoising is approximately equivalent to linear filtering; inversely, it is equivalent to global TV-denoising when a is large (see paragraph at the end of section 4.2). For intermediate values of a, local TV-denoising may achieve an interesting compromise between these two extreme behaviors, as illustrated in Figure 5 on a synthetic triangle signal. Indeed, the intermediate local TV-denoising model is able to combine a good restoration of intensity gaps (edges) with a regular smoothing of the regular parts while limiting the cropping of the extrema. It also avoids the staircasing artifact occurring in global TV-denoising, as we shall see more precisely in section 6.3.

From a statistical viewpoint, this compromise can be seen as a bias-variance trade-off with respect to the locality parameter $a \in (0, \infty)$, the regularization parameter λ being fixed.

Indeed, if ε is a (white) noise process, the mean-square error made by local TV-denoising can be classically decomposed into

$$\mathbb{E}_\varepsilon \|T_{\omega^a,\lambda}(u+\varepsilon) - u\|^2 = \underbrace{\|\mathbb{E}_\varepsilon[T_{\omega^a,\lambda}(u+\varepsilon)] - u\|^2}_{\text{squared bias}} + \underbrace{\mathbb{E}_\varepsilon \|T_{\omega^a,\lambda}(u+\varepsilon) - \mathbb{E}_\varepsilon[T_{\omega^a,\lambda}(u+\varepsilon)]\|^2}_{\text{variance}}.$$

When a is close to 0, local TV-denoising is close to the identity operator, whose bias is zero but whose variance is equal to the noise variance. Conversely, when a is large, local TV-denoising is close to global TV-denoising, which has a much smaller variance but a large bias (in particular because it kills extrema, as shown in Figure 5, top right). The bias and variance terms do not seem to follow a simple law with respect to a, but in general the best compromise is found for a finite (nonzero) value of a, as illustrated in Figure 5.

6.3. Reduction of the staircasing effect. In this section we show how local TV-denoising is able to circumvent a strong drawback of global TV-denoising. Indeed, an image denoised by global TV-denoising tends to contain blocks with constant gray level, separated by intensity gaps (edges), even in what should be smooth areas. This so-called staircasing effect also occurs in other methods, such as denoising by neighborhood filters [18], and in image or video compression methods.

The staircasing effect in global TV-denoising was first reported by Dobson and Santosa [36] and used to denoise piecewise constant images. The first mathematical proofs for its existence were successively given by Nikolova [50, 51] and Ring [55] in different contexts. Theoretical research on the staircasing artifact is active, as shown by [49] and [21], for instance. Several methods have been proposed to address this issue: Chambolle and Lions [24], Blomgren et al. [15], and Levine, Chen, and Stanich [46] propose considering modifications of TV for small gradient intensities. Chan, Marquina, and Mulet [25] and Chan, Esedoglu, and Park [29] propose using higher-order terms to capture the smooth regions. In [47], the authors propose using the ROF energy in a mean-square error sense, which annihilates the staircasing effect.

The local TV-filter we propose here is naturally free from staircasing because, thanks to Proposition 5 (section 4.4.1), when v is locally flat enough, local TV-denoising is equivalent to a blur by a low-pass filter, which naturally avoids the creation of spurious edges in smooth regions. In Figure 6, we can observe in the denoised images that contrasted edges are well preserved as in global TV-denoising, while smooth regions are much more faithfully reconstructed.

Despite the fact that local TV-denoising visually improves global TV-denoising in a significant way, it still suffers from the inability of total variation to cope with textures that are not distinguished from noise and hence systematically "washed out" by TV-based methods. Even if this drawback can be ignored for some kinds of images (in particular scientific images aiming at measuring geometrical features), it puts TV-based denoising methods a step behind state-of-the-art denoising methods. However, as we shall see now, having transformed TV-denoising into a local filter opens interesting perspectives for efficient image denoising, particularly in combination with the recent NL-means algorithm [17].

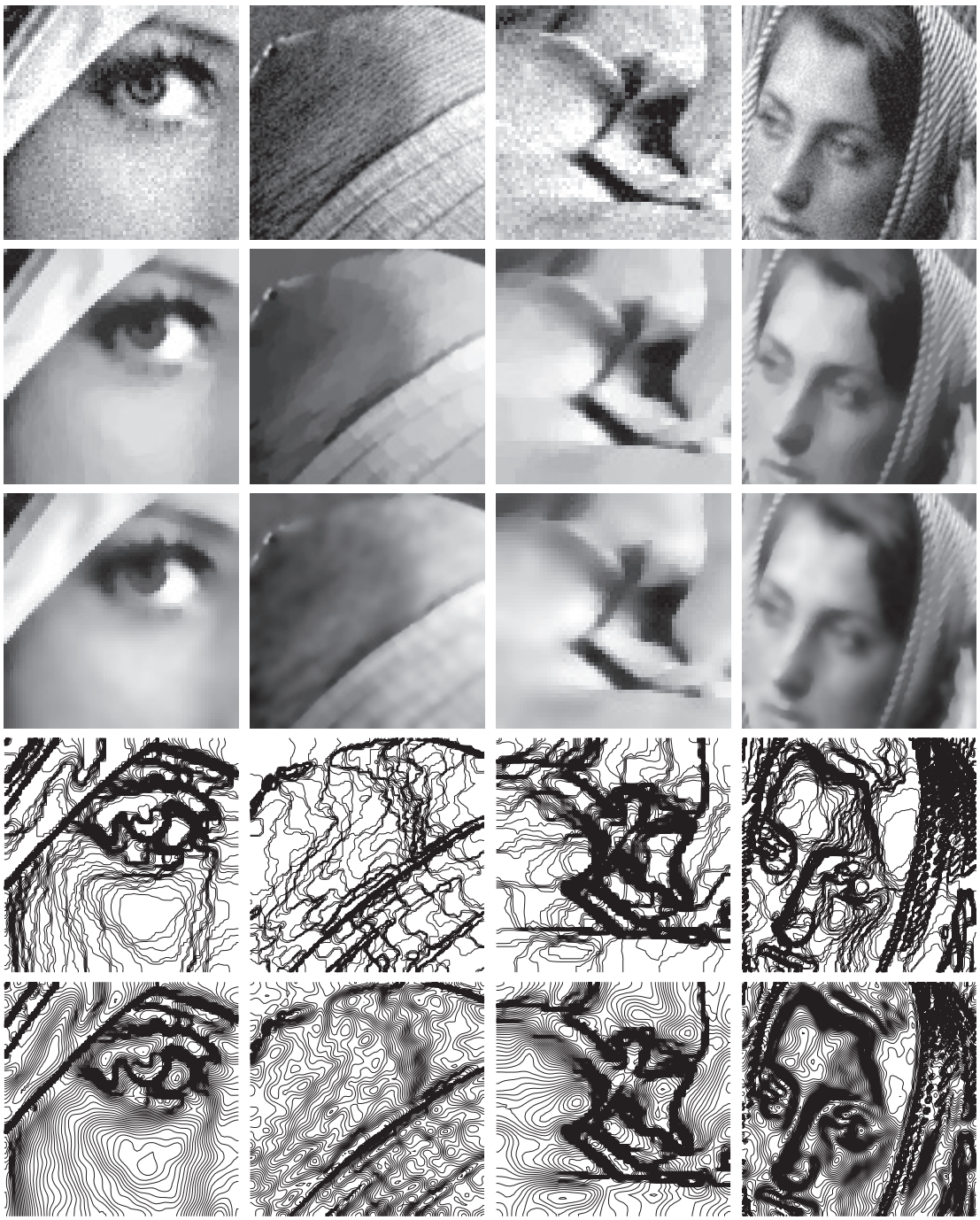


Figure 6. *Removal of the staircasing artifact. On each column, the noisy image* ($\sigma = 10$) *of the first row is denoised using global TV (second row) and local TV (third row) with* $a = 2$ *(smooth* 13×13 *window), both for* $\lambda = 40$*. As we can see, the staircasing effects (artificial edges) that appear with global TV-denoising completely disappear with local TV. This is confirmed in rows* 4 *and* 5 *(which display the level lines of the images of rows* 2 *and* 3*, respectively): instead of being artificially grouped together due to the staircasing effect (row* 4*), the level lines in smooth areas are much more regularly spaced in the case of local TV-denoising (row* 5*).*

7. Application to a hybrid TV–NL-means denoising filter. In this part, we show how the local TV-filter we studied and the NL-means denoising algorithm [17] can be combined together to build a new denoising filter that significantly improves the two methods taken separately. The NL-means algorithm, recently introduced in the continuity of the bilateral filter [62, 65] and patch-based models for inpainting or texture synthesis [31, 38], has shown how the redundancy of image patches could be used to achieve very interesting denoising performances, especially in textured areas. For this reason, its combination with TV-denoising, whose principal drawback is its inability to deal with textures, seems natural. Note that our goal here is simply to show that the combination of two very different principles (patch-based denoising and TV-regularization) can lead to interesting new denoising algorithms; finding the best possible combination would require a dedicated study that cannot be done here.

7.1. NL-means. Whereas classical local filters build an estimate of the true (that is, noise-free) gray level at a given pixel x by averaging gray levels of pixels y that are located near x (that is, for which $|y-x|$ is small), the NL-means algorithm uses pixels y that are *similar* to x, in the sense that the image patches $u(\mathcal{N}_x)$ and $u(\mathcal{N}_y)$ are similar (here $\mathcal{N}$ is a fixed patch shape, typically a 7×7 square). More precisely, an image $v\in\mathbb{R}^\Omega$ is denoised by

$$\mathrm{NLmeans}(v)(x)=\frac{\sum_{y\in\Omega}\omega_{x,y}v(y)}{\sum_{y\in\Omega}\omega_{x,y}}, \tag{54}$$

where each weight

$$\omega_{x,y}=\exp\left(-\frac{d(v(\mathcal{N}_x),v(\mathcal{N}_y))^2}{2h^2}\right) \tag{55}$$

is a similarity measure between the patches $v(\mathcal{N}_x)$ and $v(\mathcal{N}_y)$, based on the Gaussian-weighted Euclidean distance

$$d(v(\mathcal{N}_x),v(\mathcal{N}_y))=\left(\frac{\sum_{k\in\mathcal{N}}\alpha_k\,(v(x+k)-v(y+k))^2}{\sum_{k\in\mathcal{N}}\alpha_k}\right)^{\frac{1}{2}} \tag{56}$$

with $\alpha_k=e^{-|k|^2/(2a^2)}$ and $a\in(0,+\infty]$ (the case $a=\infty$ corresponds to the unweighted case). Using a weighted norm in the patch comparisons is a small improvement that permits us to give more importance to values in the center of the patch. A reasonable choice for a is $a=\frac{s-1}{4}$ for an $s\times s$ patch, so that the Gaussian weight varies in $[e^{-4},e^{-2}]\simeq[0.018,0.05]$ on the patch boundary. For computational reasons, the exploration domain of y occurring in the sums of (54) is generally restricted to a (not so large) neighborhood $\mathcal{W}_x$ of x. This makes the NL-means algorithm less "nonlocal", but brings a significant improvement in terms of computational time *and* denoising performances.

The efficiency of the NL-means method comes from the fact that in a natural image, most patches are close to several other patches (in the sense defined above). Let us be a little more precise and say that a patch $v(\mathcal{N}_y)$ is a *replica* of $v(\mathcal{N}_x)$ if they both come from the same original (noise-free) patch to which independent noises have been added. If the patch $v(\mathcal{N}_y)$ is a replica of $v(\mathcal{N}_x)$ and $\mathcal{N}_x\cap\mathcal{N}_y=\emptyset$, then we have

$$d(v(\mathcal{N}_x),v(\mathcal{N}_y))^2=2\sigma^2\frac{\sum_{k\in\mathcal{N}}\alpha_k Z_k^2}{\sum_{k\in\mathcal{N}}\alpha_k}, \tag{57}$$

15 × 15 search window $n = 6$ | 15 × 15 search window $n = 10$ | global search $n = 10$

Figure 7. *Location of rare patches. On the noisy Lena image ($\sigma = 20$), the central pixels of 11×11 patches that have fewer than n replicas are represented in black (on a light gray background showing the original* Lena *image). In the left and middle images, a local 15×15 search window is used, whereas the search for replicas is performed on the whole domain in the right image. Replicas of a patch $v(\mathcal{N}_x)$ are patches $v(\mathcal{N}_y)$ such that $d(v(\mathcal{N}_x), v(\mathcal{N}_y))^2 < \tau$ with $\tau = 2\sigma^2 \cdot 1.392$ (which corresponds to the threshold defined by (60) for 11×11 patches). As we can see, "rare patches" mostly consist of singular points like T-junctions or more generally special edge points, and occur in complex textured zones where no local autosimilarity can be found. The use of a global search window is not interesting in practice, because even if it decreases the number of rare patches (as shown in the right image), it also reduces the efficiency of NL-means denoising. A similar effect is observed with the UINTA method and is discussed in* [11].

where (Z_k) are i.i.d. random variables with distribution $\mathcal{N}(0, 1)$. In particular we have

$$\mathbb{E}\left[d(v(\mathcal{N}_x), v(\mathcal{N}_y))^2\right] = 2\sigma^2,$$

and the classical choice $h = \sigma$ ensures that most replicas have a nonnegligible weight $\omega_{x,y}$.

The main weakness of NL-means denoising is probably the way it handles "rare" patches, that is, image patches $v(\mathcal{N}_x)$ for which $d(v(\mathcal{N}_x), v(\mathcal{N}_y))^2 \gg 2\sigma^2$ for almost all pixels y (examples of such rare patches are shown in Figure 7). Due to the definition of the weights $\omega_{x,y}$ (55) and the fact that $\omega_{x,x} = 1$, the NL-means algorithm will either leave a lot of noise (for small values of h) or produce an exaggerated blur (for larger values of h). This last effect comes from the fact that the averaging is done with pixel values coming from quite different patches, leading to a "patch jittering" blur effect. In practice, this phenomenon is commonly observed on images processed with the NL-means algorithm: a significant amount of noise is left on some particular image structures such as corners, T-junctions, and rare texture patches, and several image parts suffer from a noticeable blur effect. An appropriate choice of the parameter h permits us to reduce either of these two effects, but not both simultaneously. Another consequence of this "patch jittering" blur effect is the fact that when the patch size gets larger, the NL-means filter tends to produce more blurry images (see Figure 8, right column), although intuitively, the decreasing number of similar patches should lead to a lower level of denoising, producing a noise effect.

Figure 8. *Denoising of Barbara image (detail). The classic* Barbara *image, corrupted with a Gaussian white noise with standard deviation* $\sigma = 20$ *(top left), is denoised using several methods compared in this paper. The ROF method (bottom left) does not manage to handle well the smooth and textured parts: most textured parts (scarf, basket chair) are poorly reconstructed, while a nonnegligible amount of noise remains in the originally smooth regions (cheek, hand). The NL-means method (right column, top and bottom rows) performs better, but some regions are oversmoothed (eye, basket chair) due to the* patch-jittering effect *mentioned at the end of section 7.1. This effect is particularly strong for* 11×11 *patches (top row); it can be reduced by using smaller patches (*7×7 *in bottom row), which leads to a better image but with a noticeable loss of denoising in some parts (hand, cheek). The TV-means algorithm we propose (middle column) manages to find an interesting compromise: it avoids the patch-jittering effect without reducing too much the amount of denoising in other regions, thanks to the use of local TV-denoising. The aggregated version (middle column, bottom row) yields a balanced restoration, visually nicer and significantly better than other methods in terms of the peak signal-to-noise ratio (PSNR) (see Table 1). Note that all images above were enhanced by a same affine contrast change for an easier visualization.*

7.2. A new filter: TV-means. In order to take care of rare patches, we propose changing the NL-means strategy in two ways:

(A) select only patches that could be a replica of the current patch;

(B) if the number of selected patches is too small, apply TV-regularization.

Step (A) was proposed before in [16, 32, 48, 52] and, in a different way, in [45]. Notice that the strategy (A) + (B) we propose simply consists of smoothing rare patches with a well-established method (TV-denoising) instead of using the indirect "patch-jittering" smoothing

effect. This combination of local (patch) TV-denoising and NL-means yields the following algorithm called *TV-means*.

For a given patch shape $\mathcal{N}$, a given search window $\mathcal{W}$ (as in NL-means), and a threshold τ (which will be specified afterwards), let us define, for every $\lambda \geq 0$, the set

$$\Omega(x,\lambda) = \left\{ y \in \mathcal{W}_x,\ d\left(T_\lambda(v(\mathcal{N}_x)), T_\lambda(v(\mathcal{N}_y))\right)^2 < \tau \right\}, \tag{58}$$

estimating the location of possible replicas of $v(\mathcal{N}_x)$ after TV-filtering with parameter λ. Now let $(n_\lambda)_{\lambda\geq 0}$ denote a nonincreasing sequence (this sequence controls the level of denoising and will be discussed afterwards). To every $x \in \Omega$, we associate

$$\hat{\lambda}(x) = \min\{\lambda \geq 0,\ |\Omega(x,\lambda)| \geq n_\lambda\},$$

which represents the minimal TV-filtering parameter for which $T_\lambda(v(\mathcal{N}_x))$ has enough (that is, at least n_λ) replicas. Finally, we denoise the pixel x by averaging the local TV-denoising estimates, that is,

$$\mathrm{TVmeans}(v)(x) = \frac{1}{|\Omega(x,\hat{\lambda}(x))|} \sum_{y\in\Omega(x,\hat{\lambda}(x))} T^{\mathcal{N}}_{\hat{\lambda}(x)} v(y).$$

Threshold on the patch distance. The threshold τ is chosen such that a patch $v(\mathcal{N}_x)$ and one of its replicas $v(\mathcal{N}_y)$ have a probability of 0.99 of being considered as similar. Remembering that $v(\mathcal{N}_y) - v(\mathcal{N}_x)$ is a Gaussian random variable and assuming that the patch size is large enough, we can use a result from Fisher [40] giving a central limit theorem on weighted i.i.d. random variables and use, as $(Z_x^2-1)/\sqrt{2}$ are centered and normalized i.i.d. random variables, the approximation

$$\frac{\sum_{x\in\mathcal{N}} \alpha_x \frac{Z_x^2-1}{\sqrt{2}}}{\sqrt{\sum_{x\in\mathcal{N}} \alpha_x^2}} \ \sim\ \mathcal{N}(0,1). \tag{59}$$

Hence, combining (57) and (59) yields

$$d(v(\mathcal{N}_x), v(\mathcal{N}_y))^2 \ \sim\ 2\sigma^2 \left(1 + \frac{\sqrt{2s_2}}{s_1}\mathcal{N}(0,1)\right), \quad \text{where } s_p = \sum_{k\in\mathcal{N}} \alpha_k^p \quad (p=1,2).$$

A look at the cumulative function of the normal distribution tables gives

$$\mathbb{P}(d(v(\mathcal{N}_x), v(\mathcal{N}_y))^2 < \tau) = 0.99 \iff \tau \approx 2\sigma^2 \left(1 + 2.33\frac{\sqrt{2s_2}}{s_1}\right), \tag{60}$$

which is the value of τ that we choose to define $\Omega(x,\lambda)$ (58).

Required number of patches for each scale. Here we explain how the sequence $(n_\lambda)_{\lambda\geq 0}$ (the minimum number of replicas required for a TV-filtering scale λ) can be defined in a generic way. The first term n_0 represents the minimal number of replicas to be found in order to avoid TV-filtering. It is a parameter of the TV-means algorithm that sets the balance between the

TV and NL-means combination. In practice a small value is convenient, and we chose $n_0 = 10$ in all experiments (and $n_0 = 6$ in the aggregated case presented in the next section).

Then, we set the other terms n_λ ($\lambda > 0$) in order to approximately keep the same level of denoising, for any λ; that is,

$$\forall \lambda \geq 0, \quad \frac{1}{n_0}\mathbb{E}\left[\Big(v(x) - u(x)\Big)^2\right] = \frac{1}{n_\lambda}\mathbb{E}\left[\Big(T_\lambda^{\mathcal{N}} v(x) - u(x)\Big)^2\right],$$

where u denotes the original noise-free image (recall that averaging n replicas divides the noise variance by n). Now, simulations show that the approximation

$$\mathbb{E}\left[\Big(T_\lambda^{\mathcal{N}} v(x) - u(x)\Big)^2\right] \approx \sigma^2(1 - r\lambda),$$

where $r \in [0.05, 0.1]$, usually holds for relatively small values of λ and is quite stable with respect to σ. This justifies the choice of

$$n_\lambda = n_0(1 - r\lambda),$$

with $r = 0.1$, since we do not want to oversmooth rare patches. This formula was taken in all experiments of section 7.4.

7.3. Aggregated TV-means. A very simple improvement that can be brought to the TV-means algorithm consists of taking advantage of the fact that for each pixel x, we can use the algorithm described in section 5 to estimate not only a value for the pixel x but a complete denoised patch

$$\hat{u}_x : k \in \mathcal{N} \mapsto \frac{1}{|\Omega(x, \hat{\lambda}(x))|} \sum_{y \in \Omega(x, \hat{\lambda}(x))} T_{\hat{\lambda}(x)}(v(\mathcal{N}_y))(y + k).$$

All of these estimates of $u_0(x)$ (the nonobserved noise-free image at point x), obtained for all patches containing the pixel x, can then be aggregated to define the denoised image

$$x \mapsto \frac{\sum_{k \in \mathcal{N}} \alpha_k \, \hat{u}_{x+k}(-k)}{\sum_{k \in \mathcal{N}} \alpha_k},$$

which significantly improves the denoising process. We call this variant "aggregated TV-means."

At this point, we hence have three kinds of weights: the weights arising in local TV-denoising (Definition 2), those used in the patch distance (56), and those used in the aggregation process we just mentioned. In the experiments that follow, we made the simplest choice and took all weights equal to 1 in the three cases. For local TV-denoising, since the maximum time step ensuring convergence is proportional to the minimum weight, any other choice would slow down the algorithm. For the two other weight functions (patch distance and aggregation), we observed little effect on the results when taking identical functions made of nonconstant *and nonadaptive* weights. Hence the choice of equal weights seems reasonable in a nonadaptive strategy, avoiding in particular the need for extra parameters.

However, it is important to mention here that interesting improvements could certainly be obtained by using a weight function adapted to each patch. Indeed, such adaptive weights (based on inverse variances or on Stein's unbiased risk estimate [63], for instance) have been proved to be efficient for the kind of aggregation procedure we consider here (see [44] for a recent review). The use of an adaptive weight strategy is beyond the scope of this work but could deserve further study.

7.4. Experiments. The TV-means and aggregated TV-means methods are straightforward to implement. However, it is worth mentioning that a considerable acceleration can be obtained by avoiding redundant computations of local TV-filtering. For a given patch $\mathcal{N}_x$ and a given scale λ, the computation of $T_\lambda(v(\mathcal{N}_x))$ is done at most once, and the smoothed patch is stored until it is out of range of the remaining potential patches (the image is processed by increasing line numbers, so that the pixel x cannot be part of a search window after a certain line index is reached). Using this implementation (in C language), the application of the aggregated TV-means filter takes 12 seconds on a standard PC desktop for a 256×256 image (for the patch and search window sizes specified below).

We tested the effect of TV-means denoising (in both original and aggregated variants) on the five classical images used in [45]: *Barbara*, *Lena*, *Boats*, *House*, and *Peppers*. These five images were corrupted with a white Gaussian noise (standard deviation $\sigma = 20$) and then processed with the algorithms considered above. Concerning the choice of the parameters, we tried to find, for each algorithm, a set of values that yielded good results (in terms of the peak signal-to-noise ratio (PSNR)) for all five images:

- for global TV (ROF) denoising, we used the value $\lambda = 28$;
- for NL-means, we used 7×7 patches on an 11×11 search window, a patch norm coefficient $a = \frac{s-1}{4} = 1.5$, and a weight decay $h = 18$;
- for TV-means (without aggregation), we used 11×11 patches on a 15×15 search window, with $n = 10$;
- for aggregated TV-means, we used $n = 6$ (since the aggregation process is a smoothing process, it is logical to be less demanding on n) and the other parameters as in the nonaggregated case.

Note that the "optimal" parameters (patch size and search window size) found for NL-means and TV-means significantly differ. With an equal set of parameters, TV-means would regularize the images more than NL-means, thanks to the additional TV-regularization process. Hence TV-means can be more restrictive on the choice of patches, which explains why the patch size is found to be larger for TV-means. Further, as TV-means uses 0-1 weights on the patch distances, it is able to eliminate spurious patches more easily than NL-means, which explains its ability to deal with larger search windows.

The four algorithms above are compared visually in Figure 8 and 9, where parts of the images *Barbara* and *Lena* are shown.[1] The benefit of the combination of global TV-denoising and NL-means denoising clearly appears: the TV-means method, in both variants, manages to capture the best features of its two basic components—the patch redundancy used in NL-means and the nice denoising of edges of global TV-denoising. Figure 10 more specifically compares the NL-means and aggregated TV-means algorithms using close-ups on zones that include

[1]Other examples can be found at http://www.mi.parisdescartes.fr/∼moisan/tvmeans/.

Figure 9. *Denoising of* Lena *image (detail). As in Figure 8, global TV-denoising simultaneously leaves noise (and staircasing) on the smooth parts and erases some texture (hat, feathers). The NL-means method performs better, but either leaves noticeable noise (7×7 patches) or introduces too much blur (11×11 patches). A more balanced treatment of edges, flat areas, and textures is obtained with aggregated TV-means, yielding a better global impression and a more precise recovery of textures (see, e.g., the feathers and the thin stripes of the hat).*

rare patches (see Figure 7). These zones are precisely the ones where the TV-regularization process is the most active.

We also systematically evaluated the algorithms in terms of their ability to restore the original image. In each case, we computed the PSNR between the groundtruth image u_0 and the denoised image v, defined by

$$PSNR = 10\log_{10}\left(\frac{255^2}{\sum_x \big(v(x)-u_0(x)\big)^2}\right).$$

The values are reported in Table 1. As we can observe, in terms of denoising efficiency (high PSNR), the aggregated TV-means method performs significantly better than global TV (ROF) and NL-means, and yields performances similar to the results of Kervrann and Boulanger [45] (better for *Lena* and *Boats*). Knowing that, contrary to the Kervrann–Boulanger algorithm, the method we propose does not try to optimize the search window size locally, this is a

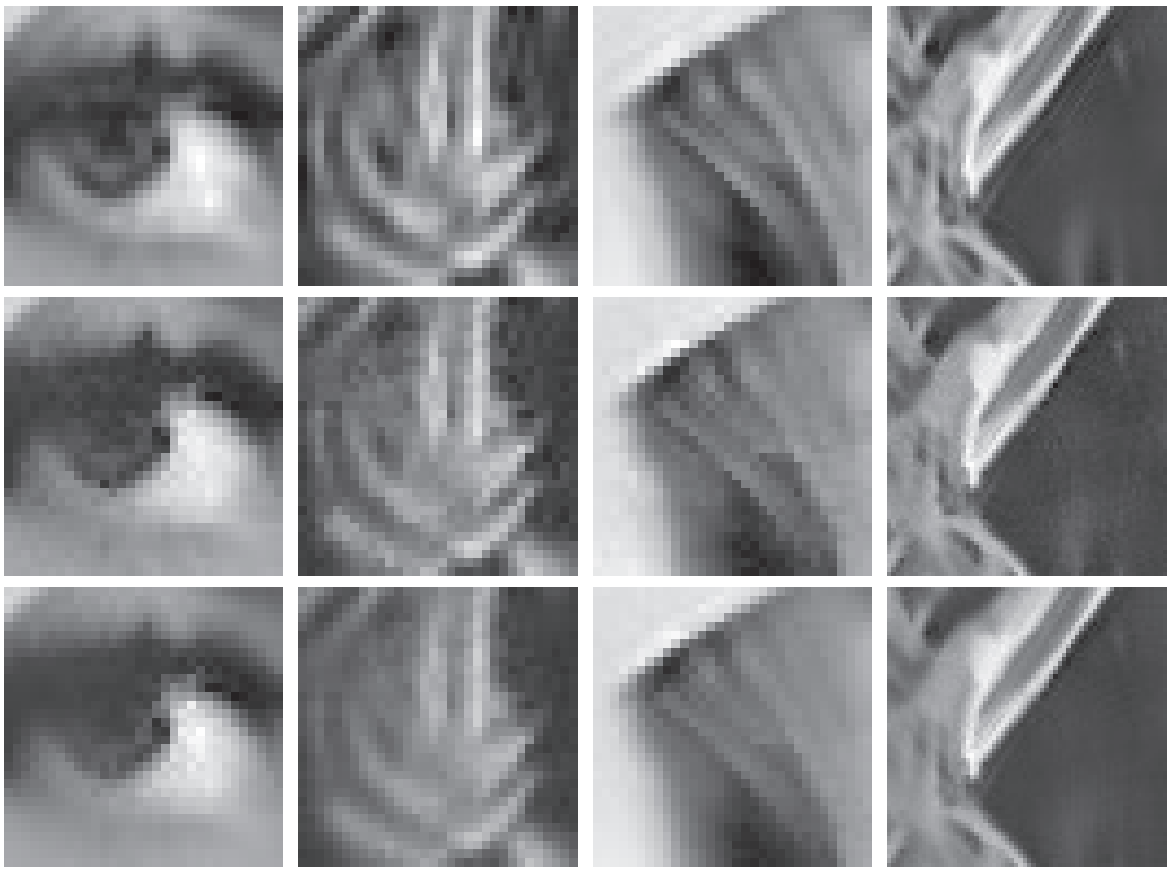


Figure 10. *Close-ups of some images in Figure 9, corresponding to zones including rare patches (see Figure 7), represented by white boxes on the noisy image in Figure 9. Row 1: aggregated TV-means (11×11 patches). Row 2: NL-means (7×7 patches). Row 3: NL-means (11×11 patches). As can also be observed in Figure 8 and 9, the TV-regularization operated by TV-means permits us to significantly improve the quality of the restoration of rare patches, compared to NL-means that either leaves noise (row 2) or introduces undesirable blur (row 3). In the two rightmost images of row 1, we can observe that TV-means restoration sometimes causes a slight ringing artifact near contrasted edges, which seems to be caused (at least, partly) by the hard thresholding used to select replicas.*

promising result, since there probably are several interesting ways to combine in a more sophisticated manner the local TV and NL-means filtering. For example, we could use a two-pass scheme as in the recent state-of-the-art denoising method BM3D [32]. Another interesting perspective would be to use the idea of shape-adaptive patches as in SA-DCT [41] or SA-PCA BM3D [33]. Local TV-denoising could be easily carried out on such patches, since the algorithm we propose is not restricted to particular domain shapes (thus we can still use uniform weights to maintain the computational efficiency). It is not clear, however, whether such a combination would be really efficient, in particular because the adaptive neighborhoods could interfere (or be redundant) with the local adaptivity of TV-denoising that we observed at the end of section 2 (see in particular the influence maps in Figure 1).

Table 1

PSNR values obtained with different denoising algorithms. We can observe that the aggregated TV-means algorithm (in bold) shows a performance similar to that of the Kervrann–Boulanger method and systematically improves both TV and NL-means filtering. Note, however, that the simple combination of TV and NL-means we propose cannot directly compete with recent sophisticated methods like BM3D (in italic).

	Barbara 512×512	Lena 512×512	Boats 512×512	House 256×256	Peppers 256×256
Noisy ($\sigma = 20$)	22.1	22.1	22.1	22.1	22.1
TV (ROF) [58]	26.69	30.89	29.21	31.22	29.62
NL-means [17]	29.59	31.50	29.32	32.05	30.12
TV-means	29.94	31.80	29.34	32.34	29.73
agg. TV-means	**30.93**	**32.48**	**30.00**	**33.10**	**30.63**
Kervrann–Boulanger [45]	30.37	32.64	30.12	32.90	30.59
BM3D [32]	*31.78*	*33.05*	*30.88*	*33.77*	*31.29*

8. Conclusion. The aim of this work was to transform the well-known ROF (TV-L^2) filter into a local filter. This transformation is interesting for several reasons and in particular because, as we showed, most image pixels have a very limited influence zone in the ROF model, so that such a local filter is expected to inherit the good behavior of the ROF model. We built a local TV-filter with an arbitrary neighborhood shape, and discovered that the introduction of an appropriate weight function (typically Gaussian) was necessary to avoid aliasing effects, which is quite intuitive considering the linear regime in which the local TV-filter falls for large values of the scale λ. Aside from interesting properties of this new local TV-filter (in particular a limiting PDE, allowing weight normalization), we established convergence conditions for the algorithms we built, inspired from Chambolle's previous works. We illustrated the interest of local TV-denoising in two directions. First, we showed that it brings an interesting bias-variance trade-off, compared to its linear regime (Gaussian filtering) and global TV-denoising. One illustration of this is the fantastic reduction of the well-known *staircasing effect*, which makes ROF-processed images look like oil paintings. Second, we used local TV-filtering to build a new filter that combines TV-denoising and NL-means into a simple but efficient denoising method called *aggregated TV-means*. This new filter brings interesting perspectives, in particular because it shows that although the ROF model is no longer a state-of-the-art denoising algorithm for most images, the idea of TV-denoising can still be used to build efficient denoising filters.

Appendix A. Proof of Proposition 6. We follow Chambolle's proof [22] and generalize it to the case of a nonrectangular and weighted domain. The fact that the domain $\mathcal{W}$ is not supposed to be rectangular is handled in the definition of the gradient and divergence operators given at the beginning of section 5.1.

(i) First, we prove that the minimizer $u \in \mathbb{R}^{\mathcal{W}}$ of $E_{\lambda,\omega,v}(u) = \|D^{1/2}(u-v)\|^2 + \lambda TV(u)$ (43) satisfies $u = v - \pi_{\frac{\lambda}{2}K}(v)$. Euler's equation for the minimization of $E_{\lambda,\omega,v}(u)$ reads as

$$2D(u-v) + \lambda \partial TV(u) \ni 0,$$

that is,

$$\frac{D(v-u)}{\lambda/2} \in \partial TV(u).$$

Denoting TV^* as the Legendre–Fenchel transform of TV, we get, thanks to the Moreau decomposition (see [56]),

$$u \in \partial TV^* \left(\frac{D(v-u)}{\lambda/2} \right).$$

Now setting $w = \frac{v-u}{\lambda/2}$, this yields

$$w - \frac{2}{\lambda} v + \frac{2}{\lambda} \partial TV^*(Dw) \ni 0, \tag{61}$$

whose left-hand term is the subdifferential of the energy $\tilde{E}_{\lambda,\omega,v}(w) = \frac{1}{2}\|w - \frac{2}{\lambda}v\|^2 + \frac{2}{\lambda}TV^*(Dw)$, and (61) implies that w minimizes $\tilde{E}_{\lambda,\omega,v}$. Moreover, if K is defined by (46), which makes sense since D is invertible, from [22] we have

$$TV^*(Dw) = \begin{cases} 0 & \text{if } w \in K, \\ +\infty & \text{else,} \end{cases}$$

which holds whatever the domain $\mathcal{W}$ as the operators ∇ and div are dual to each other. Hence the minimization of $\tilde{E}_{\lambda,\omega,v}$ amounts to projecting v onto $\frac{\lambda}{2}K$, which is closed and convex because $p \mapsto D^{-1}p$ is linear. This completes the first part of the proof.

(ii) Recalling the definition of K (46), step (1) enables us to write $w = D^{-1}\mathrm{div}\, p$, where p minimizes $\|\frac{2}{\lambda}v - D^{-1}\mathrm{div}\, p\|^2$ among all p such that $\|p\|_\infty \leq 1$. The necessary and sufficient Karush–Kuhn–Tucker conditions hold, and there exists $\alpha \in \mathbb{R}^{\mathcal{W}}$ such that

$$\forall x \in \mathcal{W}, \quad -\nabla\left(\frac{\lambda}{2} D^{-1}\mathrm{div}\, p - v\right)_x = \alpha_x p_x \tag{62}$$

with $\alpha_x \geq 0$ and $\alpha_x(|p_x|^2 - 1) = 0$ for all $x \in \mathcal{W}$. Then either $\alpha_x > 0$ and $|p_x| = 1$, or $|p_x| < 1$ and $\alpha_x = 0$. In both cases $\alpha_x = |\nabla(\frac{\lambda}{2}D^{-1}\mathrm{div}\, p - v)|_x$, and replacing this value of α_x in (62) concludes the proof of (47). ∎

Appendix B. Proof of Theorem 4. We follow the proof of [22, Theorem 3.1] in the case of a nonrectangular domain (see Appendix A).

For convenience, let L denote the linear operator $L = D^{-1/2}\mathrm{div}$, and $L^* = -\nabla D^{-1/2}$ the dual operator. Letting $X = \mathbb{R}^{\mathcal{W}}$ and $Y = (\mathbb{R}^2)^{\mathcal{W}}$, note that $L \in \mathcal{L}(Y, X)$ (linear operators mapping Y into X) and that $L^* \in \mathcal{L}(X, Y)$.

First, notice that $\|p^n\|_\infty \leq 1$ for all $n \in \mathbb{N}$, by induction on n in (49). Now, denoting $\mathbf{v} = \frac{D^{1/2}v}{\lambda/2}$, we prove that the sequence $\|Lp^n - \mathbf{v}\|^2$ is decreasing for a certain range of values of τ. Writing $\eta = (p^{n+1} - p^n)/\tau$, we have

$$\begin{aligned} &\|Lp^{n+1} - \mathbf{v}\|^2 - \|Lp^n - \mathbf{v}\|^2 \\ &\qquad = 2\tau \langle L\eta, Lp^n - \mathbf{v}\rangle + \tau^2 \|L\eta\|^2 \\ &\qquad \leq -2\tau \langle \eta, -L^*(Lp^n - \mathbf{v})\rangle_Y + \tau^2\kappa^2\|\eta\|_Y^2, \end{aligned} \tag{63}$$

where κ denotes the operator norm of L (an upper bound for κ will be given at the end of the proof). Since $\nabla(D^{-1}\mathrm{div}\, p^n - \frac{v}{\lambda/2}) = -L^*(Lp^n - \mathbf{v})$, we get from (48) that

$$\forall x \in \mathcal{W}, \quad \eta_x = -\left(L^*\left(Lp^n - \mathbf{v}\right)\right)_x - \left|\left(L^*\left(Lp^n - \mathbf{v}\right)\right)_x\right| p_x^{n+1}, \tag{64}$$

and, consequently, for any pixel x, the splitting

$$\begin{aligned}
&2\eta_x \cdot (-L^*(Lp^n - \mathbf{v}))_x \\
&\quad = |\eta_x|^2 + |(L^*(Lp^n - \mathbf{v}))_x|^2 - |\eta_x + (L^*(Lp^n - \mathbf{v}))_x|^2 \\
&\quad = |\eta_x|^2 + |(L^*(Lp^n - \mathbf{v}))_x|^2 - |(L^*(Lp^n - \mathbf{v}))_x|^2\, |p_x^{n+1}|^2,
\end{aligned}$$

implies that

$$2\,\langle \eta, -L^*(Lp^n - \mathbf{v})\rangle_Y \geq \|\eta\|_Y^2, \tag{65}$$

because $\|p^{n+1}\|_\infty \leq 1$. Thus, gathering (63) and (65) yields

$$\left\|Lp^{n+1} - \mathbf{v}\right\|^2 - \|Lp^n - \mathbf{v}\|^2 \leq -\tau\left((1 - \tau\kappa^2)\|\eta\|_Y^2\right)$$

which is negative as soon as $\tau < 1/\kappa^2$. This proves that the sequence $\|Lp^n - \mathbf{v}\|^2$ is decreasing unless $\eta = 0$, which in any case ensures that $p^{n+1} = p^n$. When $\tau = 1/\kappa^2$, the result remains true, because if $\|Lp^{n+1} - \mathbf{v}\| = \|Lp^n - \mathbf{v}\|$, then (65) is an equality, which requires that

$$\forall x \in \mathcal{W}, \quad |(L^*(Lp^n - \mathbf{v}))_x||p_x^{n+1}| = |(L^*(Lp^n - \mathbf{v}))_x|,$$

so that for a given $x \in \mathcal{W}$, either $|p_x^{n+1}| = 1$ or $(L^*(Lp^n - \mathbf{v}))_x = 0$. In both cases, $p^{n+1} = p^n$ thanks to (49).

Let m be the limit of $\|Lp^n - \mathbf{v}\|$, and let $\bar{p}$ be the limit of a converging subsequence (p^{n_k}) of (p^n). Then by (49), (p^{n_k+1}) converges to a certain $\bar{p}'$ such that

$$\forall x \in \mathcal{W}, \qquad \bar{p}'_x = \frac{\bar{p}_x + \tau(L^*(L\bar{p} - \mathbf{v}))_x}{1 + \tau|(L^*(L\bar{p} - \mathbf{v}))_x|},$$

and repeating the former computations leads, thanks to the fact that $m = \|L\bar{p} - \mathbf{v}\| = \|L\bar{p}' - \mathbf{v}\|$, to $\bar{\eta} = (\bar{p}' - \bar{p})/\tau = 0$, that is, $\bar{p} = \bar{p}'$. Thus, taking the limit ($n = n_k$, $k \to +\infty$) in (64), we get

$$\forall x \in \mathcal{W}, \quad -(L^*(L\bar{p} - \mathbf{v}))_x = |(L^*(L\bar{p} - \mathbf{v}))_x|\,\bar{p}_x,$$

which precisely characterizes the minimizer of (43), as shown in Proposition 6. Hence, $\frac{\lambda}{2}D^{-1}\mathrm{div}\,\bar{p}$ is the projection $\pi_{\frac{\lambda}{2}K}(v)$. Since this projection is unique, we deduce that the whole sequence $(\frac{\lambda}{2}D^{-1}\mathrm{div}\,p^n)_n$ tends to the desired projection, and consequently, thanks to Proposition 6, that $(v - \frac{\lambda}{2}D^{-1}\mathrm{div}\,p^n)_n$ converges towards the minimizer of (43).

Now we compute an upper bound for the norm κ of L. We have, for any η,

$$\begin{aligned}
\|L\eta\|^2 &= \|D^{-1/2}\mathrm{div}\,\eta\|^2 \\
&= \sum_{(i,j)\in\mathcal{W}} \frac{1}{\omega_{i,j}}(\eta^1_{i,j}\delta^{\mathcal{W}}_{i+1,j} - \eta^1_{i-1,j}\delta^{\mathcal{W}}_{i-1,j} + \eta^2_{i,j}\delta^{\mathcal{W}}_{i,j+1} - \eta^2_{i,j-1}\delta^{\mathcal{W}}_{i,j-1})^2 \\
&\leq \sum_{(i,j)\in\mathcal{W}} \frac{4}{\omega_{i,j}}\left[(\eta^1_{i,j}\delta^{\mathcal{W}}_{i+1,j})^2 + (\eta^1_{i-1,j}\delta^{\mathcal{W}}_{i-1,j})^2 + (\eta^2_{i,j}\delta^{\mathcal{W}}_{i,j+1})^2 + (\eta^2_{i,j-1}\delta^{\mathcal{W}}_{i,j-1})^2\right] \\
&\leq 4\sum_{i,j}{}^*\left(\frac{1}{\omega_{i,j}} + \frac{1}{\omega_{i+1,j}}\right)(\eta^1_{i,j})^2 + 4\sum_{i,j}{}^*\left(\frac{1}{\omega_{i,j}} + \frac{1}{\omega_{i,j+1}}\right)(\eta^2_{i,j})^2,
\end{aligned}$$

where the notation $\sum_{i,j}^{*} f(i,j)$ means, as in Theorem 4, that only the indices (i,j) for which $f(i,j)$ is defined are considered. This provides the upper bound

$$\kappa^2 \leq 4 \max \left\{ \max_{i,j}{}^{*} \left(\frac{1}{\omega_{i,j}} + \frac{1}{\omega_{i+1,j}} \right), \max_{i,j}{}^{*} \left(\frac{1}{\omega_{i,j}} + \frac{1}{\omega_{i,j+1}} \right) \right\},$$

so that taking $\tau \leq \tau_{\max}$ (with $\tau_{\max}$ as in (50)) yields the announced result. ∎

Acknowledgments. We thank Antonin Chambolle for fruitful discussions, and the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] G. Acosta and R. G. Durán, *An optimal Poincaré inequality in L^1 for convex domains*, Proc. Amer. Math. Soc., 132 (2004), pp. 195–202.
[2] A. Almansa, C. Ballester, V. Caselles, and G. Haro, *A TV based restoration model with local constraints*, J. Sci. Comput., 34 (2008), pp. 209–236.
[3] A. Almansa, V. Caselles, G. Haro, and B. Rougé, *Restoration and zoom of irregularly sampled, blurred, and noisy images by accurate total variation minimization with local constraints*, Multiscale Model. Simul., 5 (2006), pp. 235–272.
[4] F. Alter, V. Caselles, and A. Chambolle, *Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow*, Interfaces Free Bound., 7 (2005), pp. 29–53.
[5] F. Alter, V. Caselles, and A. Chambolle, *A characterization of convex calibrable sets in $\mathbb{R}^n$*, Math. Ann., 332 (2005), pp. 329–366.
[6] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel, *Axioms and fundamental equations of image processing*, Arch. Rational Mech. Anal., 123 (1993), pp. 199–257.
[7] F. Andreu, V. Caselles, J. I. Diaz, and J. M. Mazón, *Some qualitative properties for the total variation flow*. J. Funct. Anal., 188 (2002), pp. 516–547.
[8] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, New York, 2000.
[9] J.-F. Aujol, *Some first-order algorithms for total-variation based image restoration*, J. Math. Imaging Vision, 34 (2009), pp. 307–327.
[10] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, *Image decomposition into a bounded variation component and an oscillating component*, J. Math. Imaging Vision, 22 (2005), pp. 71–88.
[11] S. P. Awate and R. T. Whitaker, *Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, 2, 2005, pp. 44–51.
[12] G. Bellettini, V. Caselles, and M. Novaga, *The total variation flow in $\mathbb{R}^N$*, J. Differential Equations, 184 (2002), pp. 475–525.
[13] M. Bertalmio, V. Caselles, B. Rougé, and A. Solé, *A TV based restoration model with local constraints*, J. Sci. Comput., 19 (2003), pp. 95–122.
[14] J. Besag, *On the statistical analysis of dirty pictures*, J. Roy. Statist. Soc. Ser. B, 48 (1986), pp. 259–302.
[15] P. Blomgren, T. F. Chan, P. Mulet, and C. Wong, *Total variation image restoration: Numerical methods and extensions*, IEEE Trans. Image Process., 3 (1997), pp. 384–387.
[16] T. Brox and D. Cremers, *Iterated nonlocal means for texture restoration*, in Proceedings of the 1st International Conference on Scale Space and Variational Methods in Computer Vision, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 13–24.
[17] 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.
[18] A. Buades, B. Coll, and J.-M. Morel, *The staircasing effect in neighborhood filters and its solution*, IEEE Trans. Image Process., 15 (2006), pp. 1499–1595.
[19] J. Carter, *Dual Methods for Total Variation-Based Image Restoration*, Ph.D. thesis, University of California, Los Angeles, Los Angeles, CA, 2001.

[20] V. Caselles and A. Chambolle, *Anisotropic curvature-driven flow of convex sets*, Nonlinear Anal., 65 (2006), pp. 1547–1577.
[21] V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Model. Simul., 6 (2007), pp. 879–894.
[22] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.
[23] A. Chambolle, *Total variation minimization and a class of binary MRF models*, in Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Comput. Sci. 3757, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 136–152.
[24] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.
[25] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM J. Sci. Comput., 22 (2000), pp. 503–516.
[26] T. Chan, J. Shen, and H.-M. Zhou, *Total variation wavelet inpainting*, J. Math. Imaging Vision, 25 (2006), pp. 107–125.
[27] T. Chan and C. Wong, *Total variation blind deconvolution*, IEEE Trans. Image Process., 7 (1998), pp. 370–375.
[28] T. Chan, A. Yip, and F. Park, *Simultaneous total variation image inpainting and blind deconvolution*, Int. J. Imaging Syst. Technol., 15 (2005), pp. 92–102.
[29] T. F. Chan, S. Esedoglu, and F. E. Park, *Image decomposition combining staircase reduction and texture extraction*, J. Vis. Commun. Image Represent., 18 (2007), pp. 464–486.
[30] P.-L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.
[31] A. Criminisi, P. Perez, and K. Toyama, *Region filling and object removal by examplar-based image inpainting*, IEEE Trans. Image Process., 13 (2004), pp. 1200–1212.
[32] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *Image denoising by sparse* 3*D transform-domain collaborative filtering*, IEEE Trans. Image Process., 16 (2007), pp. 2080–2095.
[33] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *BM*3*D image denoising with shape-adaptive principal component analysis*, in Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'09), INRIA, 2009.
[34] J. Darbon and M. Sigelle, *A fast and exact algorithm for total variation minimization*, in Proceedings of the Second Iberian Conference on Pattern Recognition and Image Analysis, 2005, pp. 351–359.
[35] J. Darbon and M. Sigelle, *Image restoration with discrete constrained total variation. Part* I: *Fast and exact optimization*, J. Math. Imaging Vision, 26 (2006), pp. 261–276.
[36] D. C. Dobson and F. Santosa, *Recovery of blocky images from noisy and blurred data*, J. Appl. Math., 56 (1996), pp. 1181–1198.
[37] S. Durand and M. Nikolova, *Denoising of frame coefficients using* ℓ^1 *data-fidelity term and edge-preserving regularization*, Multiscale Model. Simul., 6 (2007), pp. 547–576.
[38] A. A. Efros and T. K. Leung, *Texture synthesis by non-parametric sampling*, in Proceedings of the IEEE International Conference on Computer Vision, 1999, pp. 1033–1038.
[39] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Stud. Adv. Math., CRC Press, Boca Raton, FL, 1992.
[40] E. Fisher, *A Skorohod representation and an invariance principle for sums of weighted i.i.d. random variables*, Rocky Mountain J. Math., 22 (1992), pp. 169–179.
[41] A. Foi, V. Katkovnik, and K. Egiazarian, *Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images*, IEEE Trans. Image Process., 16 (2007), pp. 1395–1411.
[42] F. Guichard and F. Malgouyres, *Total variation based interpolation*, in Proceedings of the European Signal Processing Conference, Vol. 3, Elsevier, North–Holland, Amsterdam, 1998, pp. 1741–1744.
[43] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, *Directional varying scale approximations for anisotropic signal processing*, in Proceedings of the Twelfth European Signal Processing Conference, EUSIPCO, Vienna, Austria, EURASIP, 2004, pp. 101–104.
[44] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, *From local kernel to nonlocal multiple-model image denoising*, Int. J. Comput. Vision, 86 (2010), pp. 1–32.
[45] C. Kervrann and J. Boulanger, *Local adaptivity to variable smoothness for exemplar-based image regularization and representation*, Int. J. Comput. Vision, 79 (2008), pp. 45–69.

[46] S. Levine, Y. Chen, and J. Stanich, *Image Restoration via Nonstandard Diffusion*, Technical report 04–01, Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA, 2004.
[47] C. Louchet and L. Moisan, *Total variation denoising using posterior expectation*, in Proceedings of the European Signal Processing Conference, EURASIP, 2008.
[48] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, *Non-local sparse models for image restoration*, in Proceedings of the IEEE International Conference on Computer Vision, 2009, pp. 2272–2279.
[49] F. Malgouyres, *Rank related properties for basis pursuit and total variation regularization*, Signal Process., 87 (2007), pp. 2695–2707.
[50] M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658.
[51] M. Nikolova, *Weakly constrained minimization: Application to the estimation of images and signals involving constant regions*, J. Math. Imaging Vision, 21 (2004), pp. 155–175.
[52] J. Orchard, M. Ebrahimi, and A. Wong, *Efficient nonlocal-means denoising using the SVD*, in Proceedings of the 15th IEEE International Conference on Image Processing, 2008, pp. 1732–1735.
[53] S. Osher, A. Solé, and L. Vese, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, Multiscale Model. Simul., 1 (2003), pp. 349–370.
[54] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 629–639.
[55] W. Ring, *Structural properties of solutions to total variation regularization problems*, M2AN Math. Model. Numer. Anal., 34 (2000), pp. 799–810.
[56] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[57] L. Rudin and S. Osher, *Total variation based image restoration with free local constraints*, in Proceedings of the IEEE International Conference on Image Processing, IEEE Trans. Image Process., 1 (1994), pp. 31–35.
[58] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[59] S. M. Smith and J. M. Brady, *SUSAN—A new approach to low level image processing*, Int. J. Comput. Vision, 23 (1997), pp. 45–78.
[60] D. M. Strong and T. F. Chan, *Exact Solutions to Total Variation Regularization Problems*, UCLA CAM Report 96–41, UCLA, Los Angeles, CA, 1996.
[61] H. Takeda, S. Farsiu, and P. Milanfar, *Kernel regression for image processing and reconstruction*, IEEE Trans. Image Process., 16 (2007), pp. 349–366.
[62] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, in Proceedings of the Sixth International Conference on Computer Vision, IEEE Computer Society, Washington, DC, 1998, pp. 839–846.
[63] D. Van de Ville and M. Kocher, *SURE based non-local means*, IEEE Signal Process. Lett., 16 (2009), pp. 973–976.
[64] P. Weiss, L. Blanc-Féraud, and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM J. Sci. Comput., 31 (2009), pp. 2047–2080.
[65] L. Yaroslavsky, *Digital Picture Processing. An Introduction*, Springer-Verlag, Berlin, 1985.