Multiplicative Noise Removal Using L1 Fidelity on Frame Coefficients

Sylvain Durand, Jalal M. Fadili, Mila Nikolova

Multiplicative Noise Removal Using L1 Fidelity on Frame Coefficients

Durand S.*, Fadili J.† and Nikolova M.◇

* M.A.P. 5, Université René Descartes (Paris V), 45 rue des Saint Pères, 75270 Paris Cedex 06, France
email: Sylvain.Durand@mi.parisdescartes.fr
† GREYC CNRS-ENSICAEN-Universié de Caen 6, Bd Maréchal Juin 14050 Caen Cedex, France
email: Jalal.Fadili@greyc.ensicaen.fr
◇ CMLA, ENS Cachan, CNRS, PRES UniverSud, 61 Av. President Wilson, 94230 Cachan, France
email: Mila.Nikolova@cmla.ens-cachan.fr
(*alphabetical order of the authors*)

Abstract

We address the denoising of images contaminated with multiplicative noise, e.g. speckle noise. Classical ways to solve such problems are filtering, statistical (Bayesian) methods, variational methods, and methods that convert the multiplicative noise into additive noise (using a logarithmic function), shrinkage of the coefficients of the log-image data in a wavelet basis or in a frame, and transform back the result using an exponential function.

We propose a method composed of several stages: we use the log-image data and apply a reasonable under-optimal hard-thresholding on its curvelet transform; then we apply a variational method where we minimize a specialized criterion composed of an ℓ^1 data-fitting to the thresholded coefficients and a Total Variation regularization (TV) term in the image domain; the restored image is an exponential of the obtained minimizer, weighted in a way that the mean of the original image is preserved. Our restored images combine the advantages of shrinkage and variational methods and avoid their main drawbacks. For the minimization stage, we propose a properly adapted fast minimization scheme based on Douglas-Rachford splitting. The existence of a minimizer of our specialized criterion being proven, we demonstrate the convergence of the minimization scheme. The obtained numerical results outperform the main alternative methods.

1 Introduction

In various active imaging systems, such as synthetic aperture radar, laser or ultrasound imaging, the data representing the underlying (unknown image) $S_0 : \Omega \to \mathrm{R}_+$, $\Omega \subset \mathrm{R}^2$, are corrupted with multiplicative noise. It is well known that such a noise severely degrades the image (see Fig. 2(a)). In order to increase the chance of restoring a cleaner image, several independent measurements for the same image are realized, thus yielding a set of data:

$$S_k = S_0\, \eta_k + n_k, \quad \forall k \in \{1, \cdots, K\}, \tag{1}$$

where $\eta_k : \Omega \to \mathrm{R}_+$, and n_k represent the multiplicative and the additive noise relevant to each measurement k. Usually, n_k is white Gaussian noise. A commonly used and realistic model for the distribution of η_k is the one-sided exponential distribution:

$$\eta_k \quad : \quad \mathrm{pdf}(\eta_k) = \mu\, e^{-\mu \eta_k}\, \mathbb{1}_{\mathrm{R}_+}(\eta_k);$$

the latter is plotted in Fig. 1(a). Let us remind that $1/\mu$ is both the mean and the standard deviation of this distribution. The usual practice is to take an average of the set of all measurements—such an image can be seen in (see Fig. 2(b)). Noticing that $\dfrac{1}{K}\sum_{k=1}^{K} n_k \approx 0$, the data production model reads

$$S = \frac{1}{K}\sum_{k=1}^{K} S_k = S_0\,\frac{1}{K}\sum_{k=1}^{K}\eta_k = S_0\,\eta, \tag{2}$$

see e.g. [4,60,64] and many other references. A reasonable assumption is that all η_k are independent and share the same mean μ. Then the resultant mean of the multiplicative noise η in (2) is known to follow a Gamma distribution,

$$\eta = \frac{1}{K}\sum_{k=1}^{K}\eta_k : \quad \mathrm{pdf}(\eta) = \left(\frac{K}{\mu}\right)^K \frac{\eta^{K-1}}{\Gamma(K)}\exp\left(-\frac{K\eta}{\mu}\right), \tag{3}$$

where Γ is the usual Gamma-function and since K is integer, $\Gamma(K) = (K-1)!$. Its mean is again μ and its standard deviation is μ/K. It is shown in Fig. 1(b).

Various adaptive filters for the restoration of images contaminated with multiplicative noise have been proposed in the past, e.g. see [37,65] and the numerous references therein. It can already been seen that filtering methods work well basically when the noise is moderate or weak, i.e. when K is large. Bayesian or variational methods have been proposed as well; one can consult for instance [10,36,52,62] and the references cited therein.

A large variety of methods—see e.g. [5,34], more references are given in § 1.1—rely on the conversion of the multiplicative noise into additive noise using

$$v = \log S = \log S_0 + \log\eta = u_0 + n. \tag{4}$$

In this case the probability density function of n reads (see Fig. 1(c)):

$$n = \log\eta : \quad \mathrm{pdf}(n) = \left(\frac{K}{\mu}\right)^K \frac{1}{\Gamma(K)}\exp K(n-\mu e^n)\ . \tag{5}$$

One can prove that

$$\mathbb{E}\,[n] \;=\; \psi_0(K) - \log K\ , \tag{6}$$

$$\mathrm{Var}\,[n] \;=\; \psi_1(K), \tag{7}$$

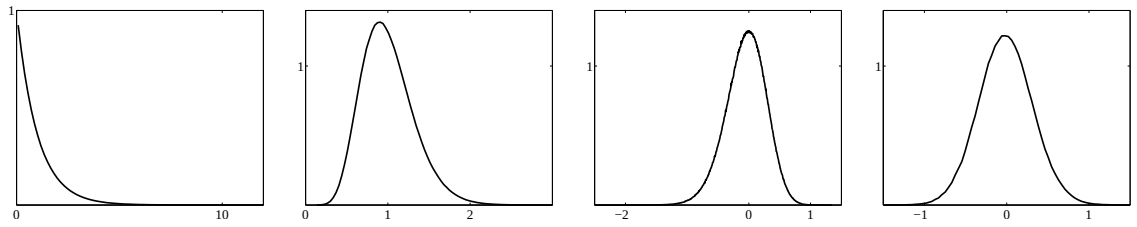
where

$$\psi_k(z) = \left(\frac{d}{dz}\right)^{k+1}\log\Gamma(z) \tag{8}$$

is the polygamma function [2].

Classical SAR modeling—see [59,60] and many other references—correspond to $\mu = 1$ in (3). Then (3) and (5) boil down to

$$\begin{aligned}\mathrm{pdf}(\eta) &= \frac{K^K\eta^{K-1}e^{-K\eta}}{(K-1)!}\ ,\\ \mathrm{pdf}(n) &= \frac{K^K e^{K(n-e^n)}}{(K-1)!}\ .\end{aligned} \tag{9}$$



(a) η_k (b) $\eta = \frac{1}{K}\sum_{k=1}^{K} \eta_k$ (c) $n = \log \eta$ (d) Wn

Figure 1: Noise distributions.

1.1 Multiscale shrinkage for the log-data

Many authors—see [4,6,34,49,64] and the references given there—focus on restoring the log-data as given in (4). The common strategy is to decompose the log-data into some multiscale frame for $L^2(\mathrm{R}^2)$, say $\{\langle w_i \rangle : i \in I\}$:

$$y = Wv = Wu_0 + Wn, \tag{10}$$

where W is the corresponding frame analysis operator, i.e. $(Wv)[i] = \langle v, w_i \rangle$, $\forall i \in I$. The rationale is that the noise Wn in y is nearly Gaussian—as seen in Fig. 1(d)—and justified by the Central Limit Theorem. The obtained coefficients y have been considered in different frameworks in the literature. In a general way, coefficients are restored using shrinkage estimators using a symmetric function $\mathcal{T} : \mathrm{R} \to \mathrm{R}$, thus yielding

$$y_{\mathcal{T}}[i] = \mathcal{T}\big((Wv)[i]\big), \quad \forall i \in I. \tag{11}$$

Following [27], various shrinkage estimators $\mathcal{T}$ have been explored in the literature, [7, 13, 28, 43, 55, 63]; see § 2.1 for more details on shrinkage methods. Shrinkage functions specially designed for multiplicative noise were proposed e.g. in [4, 6, 64].

Let $\widetilde{W}$ be a left inverse of W, giving rise to the dual frame $\{\widetilde{w}_i : i \in I\}$. Then a denoised log-image $v_{\mathcal{T}}$ is generated by expanding the shrunk coefficients $y_{\mathcal{T}}$ in the dual frame:

$$v_{\mathcal{T}} = \sum_{i \in I} \mathcal{T}((Wv)[i])\, \widetilde{w}_i = \sum_{i \in I} \mathcal{T}(y[i])\, \widetilde{w}_i. \tag{12}$$

Then the sought-after image is of the form $S_{\mathcal{T}} = \exp v_{\mathcal{T}}$.

1.2 Our approach and organization of the paper

We first apply (4) and then consider a tight-frame transform of the log-data. Our method to restore the log-image is presented in section 2. It is based on the minimization of a criterion composed of an ℓ^1-fitting to the (suboptimally) hard-thresholded frame coefficients and a Total Variation (TV) regularization in the image domain. This method uses some ideas from a previous work of some of the authors [29]. The minimization scheme to compute the log-restored image, explained in section 3, uses a Douglas-Rachford splitting specially adapted to our criterion. Restoring the sought-after image from the restored log-image requires a bias correction which is presented in section 4. The resultant algorithm to remove the multiplicative noise is provided in section 5. Various experiments are presented in section 6. Concluding remarks are given in section 7.

2 Restoration of the log-image

In this section we consider how to restore a good log-image given data $v : \Omega \to \mathrm{R}$ obtained according to (4). We focus basically on methods which, for a given preprocessed data set, lead to convex optimization problems. Below we comment only variational methods and shrinkage estimators since they underly the method proposed in this paper.

2.1 Drawbacks of shrinkage restoration and variational methods

Shrinkage restoration. The major problems with shrinkage denoising methods, as sketched in (11)-(12), is that shrinking large coefficients entails an erosion of the spiky image features, while shrinking small coefficients towards zero yields Gibbs-like oscillations in the vicinity of edges and a loss of texture information. On the other hand, if shrinkage is not sufficiently strong, some coefficients bearing mainly noise will remain almost unchanged—we call such coefficients *outliers*—and (12) suggests they generate artifacts with the shape of the functions $\widetilde{w}_i$ of the frame. A well instructive illustration can be seen in Fig. 2(b-h). Several improvements, such as translation invariant thresholding [22] and block thresholding [21], were brought to shrinkage methods in order to alleviate these artifacts. Results obtained using the latter method are presented in Figs. 3(c), 4(d) and 5(d) in Section 6. Another inherent difficulty comes from the fact that coefficients between different scales are not independent, as usually assumed, see e.g. [8, 13, 43, 54].

Variational methods. In variational methods, the restored function is defined as the minimizer of a criterion $\mathcal{F}_v$ which balances trade-off between closeness to data and regularity constraints,

$$\mathcal{F}_v(u) = \rho \int_\Omega \psi\big(u(t), v(t)\big)\, dt + \int_\Omega \varphi(|\nabla u(t)|)\, dt, \tag{13}$$

where $\psi : \mathrm{R}_+ \to \mathrm{R}_+$ helps to measure closeness to data, ∇ stands for gradient (possibly in a distributional sense), $\varphi : \mathrm{R}_+ \to \mathrm{R}_+$ is called a potential function and $\rho > 0$ is a parameter. A classical choice for ψ is $\psi(u(t), v(t)) = \big(u(t) - v(t)\big)^2$ which assumes that the noise n in (4) is white, Gaussian and centered. Given the actual distribution of the noise in (9) and Fig. 1(c), this may seem hazardous; we reconsider this choice in (15). A reasonable choice is to use the log-likelihood of n according to (9) and this was involved in the criterion proposed in [36]—see (16) at the end of this paragraph.

Let us come to the potential function φ in the regularization term. In their pioneering work, Tikhonov and Arsenin [56] considered $\varphi(t) = t^2$; however it is well known that this choice for φ leads to smooth images with flattened edges. Based on a fine analysis of the minimizers of $\mathcal{F}_v$ as solutions of PDE's on Ω, Rudin, Osher and Fatemi [53] exhibited that $\varphi(|\nabla u(t)|) = \|\nabla u(t)\|_2$, where $\|.\|_2$ is the L^2-norm, leads to images involving edges. The resultant regularization term is known as Total Variation (TV). However, whatever smooth data-fitting is chosen, this regularization yields images containing numerous constant regions (the well known stair-casing effect), so that textures and fine details are removed, see [46]. The method in [10] is of this kind and operates only on the image domain; the fitting term is derived from (3) and the criterion reads

$$\mathcal{F}_S(\Sigma) = \rho \int \left(\log \Sigma(t) + \frac{S(t)}{\Sigma(t)} \right) dt + \|\Sigma\|_{\mathrm{TV}}, \tag{14}$$

where ρ depends on K. The denoised image $\hat{S}_0 = \arg\min_{\Sigma} \mathcal{F}_S$ exhibit constant regions, as seen in Figs. 4(e) and 5(e) in Section 6. We also tried to first restore the log-image $\hat{u}$ by minimizing

$$F_v(u) = \rho\|u - v\|^2 + \|u\|_{\mathrm{TV}} \tag{15}$$

and the sought after image is of the form $\hat{S}_0 = B\exp(\hat{u})$ where B stands for the bias correction explained in section 4. Because of the exponential transform, there is no stair-casing, but some outliers remain visible—see Figs. 4(c) and 5(c); nevertheless, the overall result is very reasonable. The result of [53] was at the origin of a large amount of papers dedicated to constructing edge-preserving convex potential functions, see e.g. [3, 20, 61], and for a recent overview, [11]. Even though smoothness at the origin alleviates stair-casing, a systematic drawback of the images restored using all these functions φ is that the amplitude of edges is underestimated—see e.g. [47]. This is particularly annoying if the sought-after function has neat edges or spiky areas since the later are subjected to erosion. A very recent method proposed in [36] restores the discrete log-image using the log-likelihood of (9) and a regularized TV; more precisely,

$$\mathcal{F}_v(u, w) = \sum_i \left(u[i] + S[i]e^{-u[i]}\right) + \rho_0\|u - w\|^2 + \rho\|w\|_{\mathrm{TV}}, \tag{16}$$

where the denoised log-image $\hat{u}$ is obtained using alternate minimization on u and w. The TV term here is regularized via $\|u-w\|_2^2$ and the resultant denoised image is given by $\hat{S}_0 = \exp(\hat{w})$. The results present some improvement with respect to the method proposed in [10], at the expense of two regularization parameters (ρ and ρ_0) and twho stopping rules for each one of the minimization steps.

2.2 Hybrid methods

Hybrid methods [14, 16, 19, 23, 30, 33, 40, 41] combine the information contained in the large coefficients $y[i]$, obtained according to (10), with pertinent priors directly on the log-image u.

Remark 1 *Such a framework is particularly favorable in our case since the noise $Wn[i],\ i \in I$ in the coefficients $y[i],\ i \in I$, have a nearly Gaussian distribution—see Fig. 1(d).*

Although based on different motives, hybrid methods amount to define the restored function $\hat{u}$ as

$$\text{minimize}\ \ \Phi(u)$$

$$\text{subject to}\ \ \hat{u} \in \{u :\ |(W(u - v))\,[i]| \leq \mu_i,\ \forall i \in I\}\,.$$

If the use of an edge-preserving regularization, such as TV for Φ is a pertinent choice, the strategy for the selection of parameters $\{\mu_i\}_{i\in J}$ is more tricky. This choice must take into account the magnitude of the relevant data coefficient $y[i]$. However, deciding on the value of μ_i based solely on $y[i]$, as done in these papers, is too rigid since there are either correct data coefficients that incur smoothing ($\mu_i > 0$), or noisy coefficients that are left unchanged ($\mu_i = 0$). A way to alleviate this situation is to determine $(\mu_i)_{i\in I}$ based both on the data and on a prior regularization term. Following [44,45], this objective is carried out by defining restored coefficients $\hat{x}$ to minimize the non-smooth objective function, as explained below.

2.3 A specialized hybrid criterion

Given the log-data v obtained according to (4), we first apply a frame transform as in (10) to get $y = Wv = Wu_0 + Wn$. We systematically denote by $\hat{x}$ the denoised coefficients. The noise contained in the i-th datum reads $\langle n, w_i \rangle$ whose distribution is displayed in Fig. 1(d). The low frequency approximation coefficients carry important information about the image. In other words, when w_i is low frequency, then $\langle n, w_i \rangle$ has a better SNR than other coefficients. Therefore, as usual, a good choice is to keep them intact at this preprocessing stage. Let $I_* \subset I$ denote the subset of all such elements of the frame. Then we apply a hard-thresholding to all coefficients except those contained in I_*

$$y_{\mathcal{T}_{\rm H}}[i] \stackrel{\rm def}{=} \mathcal{T}_{\rm H}\big(y[i]\big), \quad \forall i \in I \setminus I_*, \tag{17}$$

where the hard-thresholding operator $\mathcal{T}_{\rm H}$ reads [27]

$$\mathcal{T}_{\rm H}(t) = \begin{cases} 0 & \text{if } |t| \leq T, \\ t & \text{otherwise.} \end{cases} \tag{18}$$

The resultant set of coefficients is systematically denoted by $y_{\mathcal{T}_{\rm H}}$. We choose an *underoptimal* threshold T in order to preserve as much as possible the information relevant to edges and to textures, an important part of which is contained in the small coefficients. Let us consider

$$v_{\mathcal{T}_{\rm H}} = \sum_{i \in I_1} Wv[i]\, \widetilde{w}_i, \tag{19}$$

where

$$I_1 = \{i \in I \setminus I_* : \, |y[i]| > T\}. \tag{20}$$

The image $v_{\mathcal{T}_{\rm H}}$ contains a lot of artifacts with the shape of the $\widetilde{w}_i$ for those $y[i]$ that are noisy but above the threshold T, as well as a lot of information about the fine details in the original log-image u_0. In all cases, whatever the choice of T, the image of the form $v_{\mathcal{T}_{\rm H}}$ is unsatisfactory—see Fig. 2 (b-h).

We will restore $\hat{x}$ based on the under-thresholded data $y_{\mathcal{T}_{\rm H}}$. We focus on hybrid methods of the form:

$$\begin{cases} \hat{x} & = & \arg\min_x F(x) \\ \\ F(x) & = & \Psi(x, y_{\mathcal{T}_{\rm H}}) + \Phi(\widetilde{W}x), \end{cases} \tag{21}$$

where Ψ is a data-fitting term in the domain of the frame coefficients and Φ is an edge-preserving regularization term bearing the prior on the sought-after log-image $\hat{u}$. The latter sought-after log-image $\hat{u}$ reads

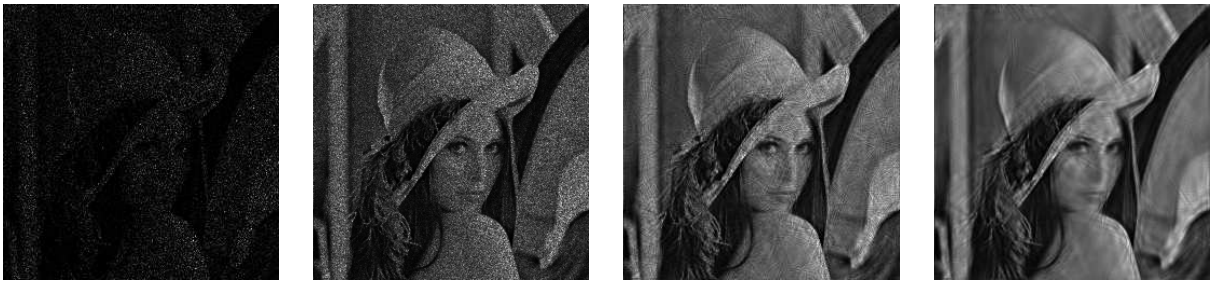
$$\hat{u} = \widetilde{W}\hat{x} \ . \tag{22}$$

Next we analyze the information content of the coefficients $y_{\mathcal{T}_{\rm H}}$ that give rise to our log-image $\hat{u}$. Let us denote

$$I_0 = I \setminus \big(I_1 \cup I_*\big) = \{i \in I \setminus I_* : \, |y[i]| \leq T\}. \tag{23}$$

We are mostly interested by the information borne by the coefficients relevant to I_0 and I_1.

(I_0) The coefficients $y[i]$ for $i \in I_0$ are usually high-frequency components which can be of the two types described below.



(a) Noisy for $K = 1$ (b) Noisy for $K = 10$ (c) $T = 2\sigma$ (d) $T = 3\sigma$

(e) $T = 4\sigma$ (f) $T = 5\sigma$ (g) $T = 6\sigma$ (h) $T = 8\sigma$

Figure 2: (a) Noisy Lena for $K = 1$. (b) Noisy Lena obtained via averaging, see (1), for $K = 10$. (c)-(h) Denoising of data v shown in (c) where the curvelet trasform of v are hard-thresholded according to (17)-(19) for different choices of T where $\sigma = \sqrt{\psi_1(K)}$ is the standard deviation of the noise n. The displayed restorations correspond to $\exp v_{\mathcal{T}_{\mathrm{H}}}$, see (19).

(a) Coefficients $y[i]$ containing essentially noise—in which case the best we can do is to keep them null, i.e. $\hat{x}[i] = y[i]$;

(b) Coefficients $y[i]$ which correspond to edges and other details in u_0. Since $y[i]$ is difficult to distinguish from the noise, the relevant $\hat{x}[i]$ should be restored using the edge-preserving prior conveyed by Φ. Notice that a careful restoration must find a nonzero $\hat{x}[i]$, since otherwise $\hat{x}[i] = 0$ would generate Gibbs-like oscillations in $\hat{u}$.

(I_1) The coefficients $y[i]$ for $i \in I_1$ are of the following two types:

(a) Large coefficients which carry the main features of the sought-after function. They verify $y[i] \approx \langle w_i, u_0 \rangle$ and can be kept intact.

(b) Coefficients which are highly contaminated by noise, characterized by $|y[i]| \gg |\langle w_i, u_0 \rangle|$. We call them *outliers* because if we had $\hat{x}[i] = y[i]$, then $\hat{u}$ would contain an artifact with the shape of $\widetilde{w}_i$ since by (19) we get $v_{\mathcal{T}_{\mathrm{H}}} = \sum_{j \setminus i} \hat{x}[j]\widetilde{w}_j + y[i]\widetilde{w}_i$. Instead, $\hat{x}[i]$ must be restored according to the prior Φ.

This analysis clearly defines the goals that the minimizer $\hat{x}$ of F_y is expected to achieve. In particular, $\hat{x}$ must involve an implicit classification between coefficients that fit to $y_{\mathcal{T}_{\mathrm{H}}}$ exactly and coefficients that are restored according to the prior term Φ. In short, restored coefficients have to fit $y_{\mathcal{T}_{\mathrm{H}}}$ exactly if they are in accordance with the regularization term Φ and have to be restored via the later term otherwise. Since [44,45] we know that criteria F_y where Ψ is non-smooth at the origin (e.g. ℓ^1) can satisfy $\hat{x}[i] = y_{\mathcal{T}_{\mathrm{H}}}[i]$

for coefficients that are in accordance with the prior Φ, while the others coefficients are restored according to Φ, see also [29]. For these reasons, we focus on a criterion on the form

$$F_y(x) = \Psi(x) + \Phi(x) \tag{24}$$

where

$$\Psi(x) = \sum_{i\in I} \lambda_i \big|(x - y_{\mathcal{T}_{\rm H}})[i]\big| = \sum_{i\in I_1\cup I_*} \lambda_i \,|(x-y)[i]| + \sum_{i\in I_0} \lambda_i \,|x[i]|\,, \tag{25}$$

$$\Phi(x) = \int_\Omega |\nabla \widetilde{W} x|\, ds = \|\widetilde{W} x\|_{\rm TV}. \tag{26}$$

Note that in (24), as well as in what follows, we write F_y in place of $F_{y_{\mathcal{T}_{\rm H}}}$ in order to simplify the notations.

In the pre-processing step (18) we would not recommend the use of a shrinkage function other than $\mathcal{T}_{\rm H}$ since it will alter *all* the data coefficients $y_{\mathcal{T}}$, without restoring them faithfully. In contrast, we base our restoration on data $y_{\mathcal{T}_{\rm H}}$ where all non-thresholded coefficients keep the original information on the sought-after image.

The theorem stated next ensures the existence of a minimizer for F_y as defined in (24) and (25)-(26). Its proof can be found in [29].

Theorem 1 [29] *For $y \in \ell^2(I)$ and $T > 0$ given, consider F_y as defined in (24), where $\Omega \in \mathrm{R}^2$ is open, bounded and its boundary $\partial\Omega$ is Lipschitz. Suppose that*

1. *$\{w_i\}_{i\in I}$ is a frame of $L^2(\Omega)$ and the operator $\widetilde{W}$ is the pseudo-inverse of W;*
2. *$\lambda_{\min} = \min_{i\in I} \lambda_i > 0$.*

Then F_y has a minimizer in $\ell^2(I)$.

Let us remind that the minimizer of F_y is not necessarily unique. Given y, denote

$$G_y \stackrel{\rm def}{=} \big\{\hat{x} \in \ell^2(I) \;:\; F_y(\hat{x}) = \min_{x\in\ell^2(I)} F_y(x)\big\}. \tag{27}$$

Hopefully, for every sample of the preprocessed data $y_{\mathcal{T}_{\rm H}}$, the set G_y is convex and corresponds to images $\widetilde{W}\hat{x}$ which are very similar since they share the same level lines. The theorem below confirms this assertion and is proven in [29].

Theorem 2 [29] *If $\hat{x}_1$ and $\hat{x}_2$ are two minimizers of F_y (i.e. $\hat{x}_1 \in G_y$ and $\hat{x}_2 \in G_y$), then*

$$\nabla\widetilde{W}\hat{x}_1 \propto \nabla\widetilde{W}\hat{x}_2, \qquad \text{a.e. on } \Omega.$$

In other words, $\widetilde{W}\hat{x}_1$ and $\widetilde{W}\hat{x}_2$ have the same level lines.

In words, images $\widetilde{W}\hat{x}_1$ and $\widetilde{W}\hat{x}_2$ are obtained one from another by a local change of contrast which is usually invisible for to the naked eye.

Some orientations for the choice of λ_i were investigated in [29]. If $i \in I_1$, the parameter λ_i should be close to, but below the upper bound $\|\widetilde{w}_i\|_{\mathrm{TV}}$, since above this bound, the coefficients $y[i]$ cannot be changed. For $i \in I_0$, a reasonable choice is

$$\lambda_i = \max_{k \neq i} \left| \int_\Omega (\nabla \widetilde{w}_i)^T \, \frac{\nabla \widetilde{w}_k}{|\nabla \widetilde{w}_k|} \, ds \right| ,$$

where $.^T$ stands for transposed. If λ_i is below this bound, some neighboring outliers might not be properly removed although Gibbs oscillations are better reduced. Another important remark is that, for some multiscale transforms, the bounds discussed above are constant. In the proposed model, we use only two values for λ_i, depending only on the set I_ϵ the index i belongs to.

We focus on the coefficients of a curvelets transforms of the log-data because (a) such a transform captures efficiently the main features of the data and (b) it is a tight-frame which is helpful for the subsequent numerical stage.

3 Minimization for the log-image

Let us rewrite the minimization problem defined in (24) and (25)-(26) in a more compact form: find $\hat{x}$ such that $F_y(\hat{x}) = \min_x F_y$ for

$$\begin{array}{l} \min_x \; F_y = \Psi + \Phi, \\ \text{where} \quad \left\{ \begin{array}{rcl} \Psi(x) & = & \|\Lambda(x - y_{\mathcal{T}_{\mathrm{H}}})\|_1 , \quad \text{for} \;\; \Lambda = \mathrm{diag}(\lambda_i)_{i \in I}, \\ \Phi(x) & = & \|\widetilde{W} x\|_{\mathrm{TV}}. \end{array} \right. \end{array} \tag{28}$$

where λ_i are the coefficients given in (25). Clearly, Ψ and Φ are proper lower-semicontinuous convex functions, hence the same holds true for F_y. The set G_y introduced in (27) is non-empty by Theorem 1 and can be rewritten as

$$G_y = \{x \in \ell^2(I) \big| x \in (\partial F_y)^{-1}(0)\},$$

where ∂F_y stands for subdifferential operator. Minimizing F_y amounts to solving the inclusion

$$0 \in \partial F_y(x) \ ,$$

or equivalently, to finding a solution to the fixed point equation

$$x = (\mathrm{Id} + \gamma \partial F_y)^{-1}(x) \ , \tag{29}$$

where $(\mathrm{Id} + \gamma \partial F_y)^{-1}$ is the *resolvent operator* associated to ∂F_y, $\gamma > 0$ is the proximal stepsize and Id is the identity map on the Hilbert space $\ell^2(I)$. The proximal schematic algorithm resulting from (29), namely

$$x^{(t+1)} = (\mathrm{Id} + \gamma \partial F_y)^{-1}(x^{(t)}),$$

is a fundamental tool for finding the root of any maximal monotone operator [31, 51], such as e.g. the subdifferential of a convex function. Since the resolvent operator $(\mathrm{Id} + \gamma \partial F_y)^{-1}$ for F_y in (28) cannot be calculated in closed-form, we focus on iterative methods.

Splitting methods do not attempt to evaluate the resolvent mapping $(\mathrm{Id} + \gamma(\partial\Psi + \partial\Phi))^{-1}$ of the combined function F_y, but instead perform a sequence of calculations involving separately the resolvent operators $(\mathrm{Id} + \gamma\partial\Psi)^{-1}$ and $(\mathrm{Id} + \gamma\partial\Phi))^{-1}$. The latter are usually easier to evaluate, and this holds true for our functionals Ψ and Φ in (28).

Splitting methods for monotone operators have numerous applications for convex optimization and monotone variational inequalities. Even though the literature is abundant, these can basically be systematized into three main classes: the forward-backward [35, 57, 58], the Douglas/Peaceman-Rachford [39], and the little-used double-backward [38, 48]. A recent theoretical overview of all these methods can be found in [24, 32]. Forward-backward can be seen as a generalization of the classical gradient projection method for constrained convex optimization, hence it inherits all its restrictions. Typically, one must assume that either Ψ or Φ is differentiable with Lipschitz continuous gradient, and the stepsizes γ must fall in a range dictated by the gradient modulus of continuity; see [26] for an excellent account. Obviously, forward-backward splitting is not adapted to our functional (29).

3.1 Specialized Douglas-Rachford splitting algorithm

The Douglas/Peaceman-Rachford family is the most general preexisting class of monotone operator splitting methods. Given a fixed scalar $\gamma > 0$, let

$$J_{\gamma\partial\Psi} \stackrel{\text{def}}{=} (\mathrm{Id} + \gamma\partial\Psi)^{-1} \quad \text{and} \quad J_{\gamma\partial\Phi} \stackrel{\text{def}}{=} (\mathrm{Id} + \gamma\partial\Phi)^{-1}. \tag{30}$$

Given a sequence $\mu_t \in (0, 2)$, this class of methods can be expressed via the recursion

$$x^{(t+1)} = \left[\left(1 - \frac{\mu_t}{2}\right)\mathrm{Id} + \frac{\mu_t}{2}(2J_{\gamma\partial\Psi} - \mathrm{Id}) \circ (2J_{\gamma\partial\Phi} - \mathrm{Id})\right] x^{(t)} \ . \tag{31}$$

Since our problem (28) admits solutions, the following result ensures that iteration (31) converges for our functional F_y.

Theorem 3 *Let $\gamma > 0$ and $\mu_t \in (0, 2)$ be such that $\sum_{t\in\mathbf{N}} \mu_t(2 - \mu_t) = +\infty$. Take $x^{(0)} \in \ell^2(I)$ and consider the sequence of iterates defined by (31). Then, $(x^{(t)})_{t\in\mathbf{N}}$ converges weakly to some point $\hat{x} \in \ell(I)$ and $J_{\gamma\partial\Phi}(\hat{x}) \in G_y$.*

This statement is a straightforward consequence of [24, Corollary 5.2]. For instance, the sequence $\mu_t = 1, \forall t \in \mathbf{N}$, satisfies the requirement of the latter theorem.

It will be convenient to introduce the reflection operator

$$\mathrm{rprox}_\varphi = 2\mathrm{prox}_\varphi - \mathrm{Id}. \tag{32}$$

where prox_φ is the proximity operator of φ according to in Definition 1. Using (35) and (32), the Douglas-Rachford iteration given in (31) becomes

$$x^{(t+1)} = \left[\left(1 - \frac{\mu_t}{2}\right)\mathrm{Id} + \frac{\mu_t}{2}\mathrm{rprox}_{\gamma\Psi} \circ \mathrm{rprox}_{\gamma\Phi}\right] x^{(t)} \ . \tag{33}$$

Below we compute the resolvent operators $J_{\gamma\partial\Psi}$ and $J_{\gamma\partial\Phi}$ with the help of Moreau proximity operators.

3.2 Proximal calculus

Proximity operators were inaugurated in [42] as a generalization of convex projection operators.

Definition 1 (Moreau [42]) *Let φ be a proper, lower-semicontinuous and convex function defined on a Hilbert space $\mathcal{H}$. Then, for every $x \in \mathcal{H}$, the function $z \mapsto \varphi(z) + \|x - z\|^2/2$, for $z \in \mathcal{H}$, achieves its infimum at a unique point denoted by $\mathrm{prox}_\varphi x$. The operator $\mathrm{prox}_\varphi : \mathcal{H} \to \mathcal{H}$ thus defined is the proximity operator of φ.*

By the minimality condition for prox_φ, it is straightforward that $\forall x, p \in \mathcal{H}$ we have

$$p = \mathrm{prox}_\varphi x \iff x - p \in \partial\varphi(p) \iff (\mathrm{Id} + \partial\varphi)^{-1} = \mathrm{prox}_\varphi. \tag{34}$$

Then (30) reads

$$J_{\gamma\partial\Psi} = \mathrm{prox}_{\gamma\Psi} \quad \text{and} \quad J_{\gamma\partial\Phi} = \mathrm{prox}_{\gamma\Phi}. \tag{35}$$

3.2.1 Proximity operator of Ψ

The proximity operator of $\gamma\Psi$ is established in the lemma stated below.

Lemma 1 *Let $x \in \ell^2(I)$. Then*

$$\mathrm{prox}_{\gamma\Psi}(x) = \Big(y_{\mathcal{T}_\mathrm{H}}[i] + {\mathcal{T}_\mathrm{S}}^{\gamma\lambda_i}\left(x[i] - y_{\mathcal{T}_\mathrm{H}}[i]\right)\Big)_{i\in I}\ , \tag{36}$$

with

$${\mathcal{T}_\mathrm{S}}^{\gamma\lambda_i}(z[i]) = \max\big\{0,\ z[i] - \gamma\lambda_i \mathrm{sign}(z[i])\big\}. \tag{37}$$

Proof. Ψ is an additive separable function in each coordinate $i \in I$. Thus, solving the proximal minimization problem of Definition 1 is also separable. For any convex function φ and $v \in \ell^2(I)$, put $\psi(.) = \varphi(. - v)$. Then

$$\begin{aligned} p = \mathrm{prox}_\psi(x) &\iff x - p \in \partial\psi(p) \\ &\iff (x - v) - (p - v) \in \partial\varphi(p - v) \\ &\iff p - v = \mathrm{prox}_\varphi(x - v) \\ &\iff p = v + \mathrm{prox}_\varphi(x - v)\ . \end{aligned}$$

For each $i \in I$, we apply this result with $v = y_{\mathcal{T}_\mathrm{H}}[i]$ and $\varphi(z[i]) = \gamma\lambda_i|z[i]|$. Noticing that $\mathrm{prox}_\varphi = {\mathcal{T}_\mathrm{S}}^{\gamma\lambda_i}$ is soft-thresholding with threshold $\gamma\lambda_i$, leads to (36). ⋄

Note that now

$$\mathrm{rprox}_{\gamma\Psi}(x) = 2\Big(y_{\mathcal{T}_\mathrm{H}}[i] + {\mathcal{T}_\mathrm{S}}^{\gamma\lambda_i}\left(x[i] - y_{\mathcal{T}_\mathrm{H}}[i]\right)\Big)_{i\in I} - x\ . \tag{38}$$

3.2.2 Proximity operator of Φ

Clearly, $\Phi(x) = \|\cdot\|_{\mathrm{TV}} \circ \widetilde{W}(x)$ is a pre-composition of the TV-norm with the linear operator $\widetilde{W}$. Computing the proximity operator of Φ for an arbitrary $\widetilde{W}$ may be intractable. We adopt the following assumptions:

(w1) $\widetilde{W} : \ell^2(I) \to L^2(\Omega)$ is surjective;

(w2) $\widetilde{W}W = \mathrm{Id}$ and $\widetilde{W} = c^{-1}W^*$ for $0 < c < \infty$, where W^* stands for the adjoint operator; note that we also have $W^*W = c\ \mathrm{Id}$;

(w3) $\widetilde{W}$ is bounded.

For any $z(t) = (z_1(t), z_2(t)) \in \mathrm{R}^2, t \in \Omega$, we set $|z(t)| = \sqrt{z_1(t)^2 + z_2(t)^2}$. Let $\mathcal{X} = L^2(\Omega) \times L^2(\Omega)$ and $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ be the inner product in $\mathcal{X}$, and $\left|\!\left|\!\left| \cdot \right|\!\right|\!\right|_p$, for $p \in [1, \infty]$ the L^p-norm on $\mathcal{X}$. We define $\overline{B}_{\infty}^{\gamma}(\mathcal{X})$ as the closed L^∞-ball of radius γ in $\mathcal{X}$,

$$\overline{B}_{\infty}^{\gamma} \stackrel{\mathrm{def}}{=} \left\{ z \in \mathcal{X} : \left|\!\left|\!\left| z \right|\!\right|\!\right|_{\infty} \leq \gamma \right\} = \Big\{ z = (z_1, z_2) \in \mathcal{X} : |z(t)| \leq \gamma, \forall t \in \Omega \Big\}, \tag{39}$$

and $P_{\overline{B}_{\infty}^{\gamma}(\mathcal{X})} : \mathcal{X} \to \overline{B}_{\infty}^{\gamma}(\mathcal{X})$ the associated projector; it is easy to check that the latter is equal to the proximity operator of the indicator function of $\overline{B}_{\infty}^{\gamma}(\mathcal{X})$. The expression for $\mathrm{prox}_{\gamma\Phi}$ is given in the next lemma while the computation scheme to solve item (ii) is stated in Lemma 3.

Lemma 2 *Let $x \in \ell^2(I)$ and $\overline{B}_{\infty}^{\gamma}(\mathcal{X})$ is as defined above.*

(i) Denoting by $\mathrm{prox}_{c^{-1}\gamma\|\cdot\|_{\mathrm{TV}}}(u)$ the proximity operator of the (c^{-1}-scaled) TV-norm, we have

$$\mathrm{prox}_{\gamma\Phi}(x) = \left(\mathrm{Id} - W \circ \left(\mathrm{Id} - \mathrm{prox}_{c^{-1}\gamma\|\cdot\|_{\mathrm{TV}}} \right) \circ \widetilde{W} \right)(x) \ ; \tag{40}$$

(ii) Furthermore,

$$\mathrm{prox}_{c^{-1}\gamma\|\cdot\|_{\mathrm{TV}}}(u) = u - P_C(u) \ , \tag{41}$$

where

$$C = \Big\{ \mathrm{div}(z) \in L^2(\Omega) \big| z \in \mathcal{C}_c^{\infty}(\Omega \times \Omega), z \in \overline{B}_{\infty}^{\gamma/c}(\mathcal{X}) \Big\}. \tag{42}$$

Proof. Since $\widetilde{W}$ is surjective, its range is $L^2(\Omega)$ which is closed. Moreover, the domain $\mathrm{dom}(\|\cdot\|_{\mathrm{TV}}) = L^2(\Omega)$ as well, so that $\mathrm{cone}\left(\mathrm{dom}\|\cdot\|_{\mathrm{TV}} - \mathrm{range}\ \widetilde{W}\right) = \{0\}$ which is a closed subspace of $L^2(\Omega)$. Reminding that $\|\cdot\|_{\mathrm{TV}}$ is lower bounded, continuous and convex, it is clear that all assumptions required in [25, Proposition 11] are satisfied. Applying the same proposition yields statement (i).

We focus next on (ii). Note that C in (42) is a closed convex subset since $\overline{B}_{\infty}^{\gamma/c}(\mathcal{X})$ is closed and convex, and div is linear; thus the projection P_C is well defined.

Let us remind that the Legendre-Fenchel (known also as the convex-conjugate) transform of a function $\varphi : \mathcal{H} \to \mathrm{R}$, where $\mathcal{H}$ is an Hilbert space, is defined by

$$\varphi^*(w) = \sup_{u \in \mathrm{dom}(\varphi)} \big\{ \langle w, u \rangle - \varphi(u) \big\},$$

and that φ^* is a closed convex function. If φ is convex, proper and lower semi-continuous, the original Moreau decomposition [42, Proposition 4.a] tells us that

$$\mathrm{prox}_{\varphi} + \mathrm{prox}_{\varphi^*} = \mathrm{Id}\ . \tag{43}$$

One can see also [26, Lemma 2.10] for an alternate proof. It is easy to check that the conjugate function of a norm is the indicator function $\imath$ of the ball of its dual norm, see e.g. [9, Eq.(2.7)]; thus

$$\left(c^{-1}\gamma\|\cdot\|_{\mathrm{TV}}\right)^*(z) = \begin{cases} 0 & \text{if } z \in C\ , \\ +\infty & \text{if } z \notin C\ , \end{cases}$$

where C is given in (42). Using Definition 1, it is straightforward that

$$\mathrm{prox}_{\left(c^{-1}\gamma\|.\|_{\mathrm{TV}}\right)^*} = P_C.$$

Identifying $c^{-1}\gamma\|.\|_{\mathrm{TV}}$ with φ and $\left(c^{-1}\gamma\|.\|_{\mathrm{TV}}\right)^*$ with φ^*, equation (43) leads to statement (ii). The proof is complete. ⋄

Note that our argument (43) for the computation of $\mathrm{prox}_{c^{-1}\gamma\|\cdot\|_{\mathrm{TV}}}(u)$ is not used in [18], which instead uses conjugates and bi-conjugates of the objective function.

Remark 2 In view of equations (41) and (42), we one can see that the term between the middle parentheses in equation (40) admits a simpler form:

$$\mathrm{Id} - \mathrm{prox}_{c^{-1}\gamma\|\cdot\|_{\mathrm{TV}}} = P_C.$$

Using (32) along with (40)-(41) we easily find that

$$\begin{aligned} \mathrm{rprox}_{\gamma\Phi}(x) &= \left(\mathrm{Id} - 2W \circ \left(\mathrm{Id} - \mathrm{prox}_{c^{-1}\gamma\|\cdot\|_{\mathrm{TV}}}\right) \circ \widetilde{W}\right)(x) \\ &= \left(\mathrm{Id} - 2W \circ P_C \circ \widetilde{W}\right)(x)\ . \end{aligned} \tag{44}$$

3.2.3 Calculation of the projection P_C in (41) in a discrete setting

In what follows, we work in the discrete setting. We consider that that W is an $M \times N$ tight frame with $M = \#I \gg N$, admitting a constant $c > 0$ such that

$$\widetilde{W}W = \mathrm{Id} \quad \text{and} \quad \widetilde{W} = c^{-1}W^T \quad (\text{ hence } \quad W^TW = c\ \mathrm{Id}).$$

(This is the discrete equivalent of assumption (w2).) We also suppose that $\widetilde{W} : \ell^2(I) \to \ell^2(\Omega)$ is surjective. Next we replace $\mathcal{X}$ by its discrete counterpart,

$$\mathcal{X} = \ell^2(\Omega) \times \ell^2(\Omega) \ \text{ where } \ \Omega \ \text{ is discrete with } \ \#\Omega = N. \tag{45}$$

We denote the discrete gradient by $\ddot{\nabla}$ and consider $\mathrm{Div} : \mathcal{X} \to \ell^2(\Omega)$ the discrete divergence defined by analogy with the continuous setting [1] as the adjoint of the gradient $\mathrm{Div} = -\ddot{\nabla}^*$; see [18].

[1] More precisely, let $u \in \ell^2(\Omega)$ be of size $m \times n$, $N = mn$. We write $(\ddot{\nabla}u)[i,j] = \big(u[i+1,j] - u[i,j],\ u[i,j+1] - u[i,j]\big)$ with boundary conditions $u[m+1,i] = u[m,i],\ \forall i$ and $u[i,n+1] = u[i,n],\ \forall i$; then for $z \in \mathcal{X}$, we have $(\mathrm{Div}(z))[i,j] = \big(z_1[i,j] - z_1[i-1,j]\big) + \big(z_2[i,j] - z_2[i,j-1]\big)$ along with $z_1[0,i] = z_1[m,i] = z_2[i,0] = z_2[i,n] = 0,\ \forall i$.

Unfortunately, the projection in (41) does not admit an explicit form. The next lemma provides an iterative scheme to compute the proximal points introduced in Lemma 2. In this discrete setting, C in (42) admits a simpler expression:

$$C = \left\{ \mathrm{Div}(z) \in \ell^2(\Omega) \;\middle|\; z \in \overline{B}_\infty^{\gamma/c}(\mathcal{X}) \right\}. \,, \tag{46}$$

where $\overline{B}_\infty^{\gamma/c}(\mathcal{X})$ is defined according to (39).

Lemma 3 *We adapt all assumptions of Lemma 2 to the new discrete setting, as explained above. Consider the forward-backward iteration*

$$z^{(t+1)} = P_{\overline{B}_\infty^1(\mathcal{X})}\left(z^{(t)} + \beta_t \ddot{\nabla} \left(\mathrm{Div}(z^{(t)}) - cu/\gamma \right) \right) \;\; \textit{for} \;\; 0 < \inf_t \beta_t \leq \sup_t \beta_t < 1/4, \tag{47}$$

where $\forall (i,j) \in \Omega$

$$P_{\overline{B}_\infty^1(\mathcal{X})}(z)[i,j] = \begin{cases} z[i,j] & \textit{if } |z[i,j]| \leq 1; \\ \dfrac{z[i,j]}{|z[i,j]|} & \textit{otherwise} \,. \end{cases}$$

Then

(i) $(z^{(t)})_{t\in\mathbf{N}}$ *converges to a point* $\hat{z} \in \overline{B}_\infty^1(\mathcal{X})$*;*

(ii) $\left(c^{-1}\gamma \mathrm{Div}(z^{(t)}) \right)_{t\in\mathbf{N}}$ *converges to* $c^{-1}\gamma\mathrm{Div}(\hat{z}) = (\mathrm{Id} - \mathrm{prox}_{c^{-1}\gamma\|\cdot\|_{\mathrm{TV}}})(u)$.

Proof. Given $u \in \ell^2(\Omega)$, the projection $\hat{w} = P_C(u)$ defined by (41) and (46) is unique and satisfies

$$\begin{aligned} \hat{w} &= \arg\min_{w\in C} \frac{1}{2}\|u-w\|^2 = \arg\min \left\{ \frac{1}{2} \left\| \frac{c}{\gamma}u - w \right\|^2 \text{ subject to } w = \mathrm{Div}(z) \text{ for } z \in \overline{B}_\infty^1(\mathcal{X}) \right\} \\ &\quad \Updownarrow \\ \hat{w} &= \mathrm{Div}(\hat{z}) \;\text{ where }\; \hat{z} = \arg\min_{z\in\overline{B}_\infty^1(\mathcal{X})} \frac{1}{2} \left\| \frac{c}{\gamma}u - \mathrm{Div}(z) \right\|^2 \end{aligned} \tag{48}$$

This problem can be solved using a projected gradient method (which is a special instance of the forward-backward splitting scheme) whose iteration is given by (47). This iteration converges weakly to a minimizer of (48)—see [24, Corollary 6.5], provided that the stepsize $\beta_t > 0$ satisfies $\sup_t \beta_t < 2/\delta^2$, where δ is the spectral norm of the div operator. It is easy to check that $\delta^2 \leq 8$—see e.g. [18].

Set

$$\omega^{(t)} = c^{-1}\gamma\mathrm{Div}(z^{(t)}), \forall t \in \mathbf{N} \;\text{ and }\; \hat{\omega} = c^{-1}\gamma\mathrm{Div}(\hat{z}).$$

Thus,

$$\begin{aligned} \left\| \omega^{(t)} - \hat{\omega} \right\|^2 &= \left(\frac{\gamma}{c} \right)^2 \left\| \mathrm{Div}(z^{(t)}) - \mathrm{Div}(\hat{z}) \right\|^2 \\ &= \left(\frac{\gamma}{c} \right)^2 \langle -\ddot{\nabla} \left(\mathrm{Div}(z^{(t)}) - \mathrm{Div}(\hat{z}) \right), z^{(t)} - \hat{z} \rangle_{\mathcal{X}} \,, \end{aligned} \tag{49}$$

where we use the fact that $-\ddot{\nabla}$ is the adjoint of Div. Let D_z denote the gradient of a scalar-valued function of z, not to be confused with the discrete gradient operator $\ddot{\nabla}$ of an image. The gradient of the

function $\frac{1}{2}\left\|cu/\gamma - \mathrm{Div}(z)\right\|^2$ with respect to z is $\mathsf{D}_z\left(\frac{1}{2}\left\|cu/\gamma - \mathrm{Div}(z)\right\|^2\right) = -\ddot{\nabla}\left(\mathrm{Div}(z) - cu/\gamma\right)$. This relation together with the Schwarz inequality applied to (49) lead to

$$\begin{aligned}
\left\|\omega^{(t)} - \hat{\omega}\right\|^2 &\leq \left(\frac{\gamma}{c}\right)^2 \left\|\ddot{\nabla}\ \mathrm{Div}(z^{(t)}) - \ddot{\nabla}\ \mathrm{Div}(\hat{z})\right\|_2 \left\|z^{(t)} - \hat{z}\right\|_2 \\
&= \left(\frac{\gamma}{c}\right)^2 \left\|\ddot{\nabla}\ \left(\mathrm{Div}(z^{(t)}) - cu/\gamma\right) - \ddot{\nabla}\ \left(\mathrm{Div}\big((\hat{z}) - cu/\gamma\right)\right\|_2 \left\|z^{(t)} - \hat{z}\right\|_2 \\
&= 0.5\left(\frac{\gamma}{c}\right)^2 \left\|\mathsf{D}_z\left(\left\|cu/\gamma - \mathrm{Div}(z^{(t)})\right\|^2\right) - \mathsf{D}_z\left(\left\|cu/\gamma - \mathrm{Div}(\hat{z})\right\|^2\right)\right\|_2 \left\|z^{(t)} - \hat{z}\right\|_2 \quad (50)
\end{aligned}$$

From [24, Theorem 6.3], we deduce that the series

$$\sum_{t\in\mathbf{N}} \left\|\mathsf{D}_z\left(\left\|cu/\gamma - \mathrm{Div}(\cdot)\right\|\right)(z^{(t)}) - \mathsf{D}_z\left(\left\|cu/\gamma - \mathrm{Div}(.)\right\|\right)(\hat{z})\right\|_2^2$$

is convergent. Inserting this property in (50) and using the fact that the sequence $(z^{(t)})_{t\in\mathbf{N}}$ is bounded (as it converges weakly with $\left\|\hat{z}\right\|_2 < \liminf_{t\to\infty}\left\|z^{(t)}\right\|_2$), it follows that $\omega^{(t)}$ converges strongly to $\hat{\omega}$. This completes the proof. ◇

The forward-backward splitting-based iteration proposed in (47) to compute the proximity operator of the TV-norm is new and different from the projection algorithm of Chambolle [18], even tough the two algorithms bear some similarities. The forward-backward splitting allows to derive a sharper upper-bound on the stepsize β_t than the one proposed in [18]—actually twice as large. Let us remind that it was observed in [18] that the bound 1/4 still works in practice. Here we prove why thus is really true.

3.3 Comments on the Douglas-Rachford scheme for F_y

A crucial property of the Douglas-Rachford splitting scheme (33) is its robustness to numerical errors that may occur when computing the proximity operators prox_Ψ and prox_Φ, see [24]. We have deliberately omitted this property in (33) for the sake of simplicity. This robustness property has important consequences: e.g. it allows us to run the forward-backward sub-recursion (47) only a few iterations to compute an approximate of the TV-norm proximity operator in the inner iterations, and the Douglas-Rachford is still guaranteed to converge provided that these numerical errors are under control. More precisely, let $a_t \in \ell^2(I)$ be an error term that models the inexact computation of $\mathrm{prox}_{\gamma\Phi}$ in (40), as the latter is obtained through (47). If the sequence of error terms $(a_t)_{t\in\mathbf{N}}$ and stepsizes $(\mu_t)_{t\in\mathbf{N}}$ defined in Theorem 3 obey $\sum_{t\in\mathbf{N}} \mu_t \left\|a_t\right\| < +\infty$, then the Douglas-Rachford algorithm (33) converges weakly [24, Corollary 6.2]. In our case, using 200 inner iterations in (47) was sufficient to satisfy this requirement.

Remark 3 Owing to the splitting framework and proximal calculus, we have shown in Lemma 2 that the bottleneck of the minimization algorithm is in the computation of the proximity-operator of the TV-norm. In fact, computing $\mathrm{prox}_{\|\cdot\|_{\mathrm{TV}}}$ amounts to solving a discrete ROF-denoising. Our forward-backward iteration is one possibility among others, and other algorithms beside [18] have been proposed to solve the discrete ROF-denoising problem. While this paper was submitted, our attention was drawn to an independent work of [12] who, using a different framework, derive an iteration similar to (47) to solve the ROF. Another parallel work of [66] propose an application of gradient projection to solving the dual problem (48). We are of course aware of max-flow/min-cut type algorithms, for instance the one in [17].

We have compared our whole denoising procedure using our implementation of $\mathrm{prox}_{\|\cdot\|_{\mathrm{TV}}}$ and the max-flow based implementation that we adapted from the code available at [1]. We obtained similar results, although the max-flow-based algorithm was faster, mainly because it uses the ℓ^1 approximation of the discrete gradient, namely $\left\|(\ddot{\nabla}u)[i,j]\right\|_1 = \left|u[i+1,j]-u[i,j]\right| + \left|u[i,j+1]-[i,j]\right|$. Let us remind that this approximation for the discrete gradient does not inherits the rotational invariance property of the L^2 norm of the usual gradient.

4 Bias correction to recover the sought-after image

Recall from (4) that $u_0 = \log S_0$ and set $\hat{u} = \widetilde{W}\hat{x}^{(N_{\mathrm{DR}})}$ as the estimator of u_0, where N_{DR} is the number of Douglas-Rachford iterations in (33). Unfortunately, the estimator $\hat{u}$ is prone to bias, i.e. $\mathbb{E}[\hat{u}] = u_0 - b_{\hat{u}}$. A problem that classically arises in statistical estimation is how to correct such a bias. More importantly is how this bias affects the estimate after applying the inverse transformation, here the exponential. Our goal is then to ensure that for the estimate $\hat{S}$ of the image, we have $\mathbb{E}\left[\hat{S}\right] = S_0$. Expanding $\hat{S}$ in the neighborhood of $\mathbb{E}[\hat{u}]$, we have

$$\begin{aligned}\mathbb{E}[\exp\hat{u}] &= \exp(\mathbb{E}[\hat{u}])(1+\mathrm{Var}[\hat{u}]/2+R_2)\\ &= S_0\exp(-b_{\hat{u}})(1+\mathrm{Var}[\hat{u}]/2+R_2)\ ,\end{aligned} \tag{51}$$

where R_2 is expectation of the Lagrange remainder in the Taylor series. One can observe that the posterior distribution of $\hat{u}$ is nearly symmetric, in which case $R_2 \approx 0$. That is, $b_{\hat{u}} \approx \log(1+\mathrm{Var}[\hat{u}]/2)$ to ensure unbiasedness. Consequently, finite sample (nearly) unbiased estimates of u_0 and S_0 are respectively $\hat{u} + \log(1+\mathrm{Var}[\hat{u}]/2)$, and $\exp(\hat{u})(1+\mathrm{Var}[\hat{u}]/2)$. $\mathrm{Var}[\hat{u}]$ can be reasonably estimated by $\psi_1(K)$, the variance of the noise n in (4) being given in (7). Thus, given the restored log-image $\hat{u}$, our restored image read:

$$\hat{S} = \exp(\hat{u})(1+\psi_1(K)/2)\ . \tag{52}$$

The authors of [64] propose a direct estimate of the bias $b_{\hat{u}}$ using the obvious argument that the noise n in the log-transformed image has a non-zero mean $\psi_0(K) - \log K$. A quick study shows that the functions $(1+\psi_1(K)/2)$ and $\exp(\log K - \psi_0(K))$ are very close for K reasonably large. Thus, the two bias corrections are equivalent. Even though the bias correction approach we propose can be used in a more general setting.

5 Full algorithm to suppress multiplicative noise

Now, piecing together Lemma 1, Lemma 2 and Theorem 3, we arrive at the multiplicative noise removal algorithm:

Task: Denoise an image S contaminated with multiplicative noise according to (2).

Parameters: The observed noisy image S, number of iterations N_{DR} (Douglas-Rachford outer iterations) and N_{FB} (Forward-Backward inner iterations), stepsizes $\mu_t \in (0,2)$, $0 < \beta_t < 1/4$ and $\gamma > 0$, tight-frame transform W and initial threshold T (e.g. $T = 2\sqrt{\psi_1(K)}$), regularization parameters $\lambda_{0,1}$ associated to the sets $I_{0,1}$.

Specific operators:

(a) $\mathcal{T}_{\mathrm{S}}{}^{\gamma\lambda_i}(z) = \Big(\max\big\{0,\ z[i] - \gamma\lambda_i \mathrm{sign}(z[i])\big\}\Big)_{i\in I}, \quad \forall z \in \mathrm{R}^{\#I}.$

(b) $P_{\overline{B}^{\,1}_{\infty}(\mathcal{X})}(z)[i,j] = \begin{cases} z[i,j] & \text{if } |z[i,j]| \le 1; \\ \dfrac{z[i,j]}{|z[i,j]|} & \text{otherwise,} \end{cases} \qquad \forall (i,j) \in \Omega.$

(c) $\ddot{\nabla}$ and Div—the discrete versions of the continuous operators ∇ and div.

(d) $\psi_1(\cdot)$ defined according to (8) (built-in Matlab function, otherwise see [50]).

Initialization:

- Compute $v = \log S$ and transform coefficients $y = Wv$. Hard-threshold y at T to get $y_{\mathcal{T}_{\mathrm{H}}}$. Choose an initial $x^{(0)}$.

Main iteration:
For $t = 1$ **to** N_{DR},

(1) Inverse curvelet transform of $x^{(t)}$ according to $u^{(t)} = \widetilde{W} x^{(t)}$.

(2) Initialize $z^{(0)}$; **For** $s = 0$ **to** $N_{\mathrm{FB}} - 1$

$$z^{(s+1)} = P_{\overline{B}^{\,1}_{\infty}(\mathcal{X})}\left(z^{(s)} + \beta_t \ddot{\nabla}\left(\mathrm{Div}(z^{(s)}) - \tfrac{c}{\gamma} u^{(t)}\right)\right).$$

(3) Set $z^{(t)} = z^{(N_{\mathrm{FB}})}$.

(4) Compute $w^{(t)} = c^{-1}\gamma\ \mathrm{Div}(z^{(t)})$.

(5) Forward curvelet transform: $\alpha^{(t)} = W w^{(t)}$.

(6) Compute $r^{(t)} = \mathrm{rprox}_{\gamma\Phi}(x^{(t)}) = x^{(t)} - 2\alpha^{(t)}$.

(7) By (38) compute $q^{(t)} = \left(\mathrm{rprox}_{\gamma\Psi} \circ \mathrm{rprox}_{\gamma\Phi}\right) x^{(t)} = 2\left(y_{\mathcal{T}_{\mathrm{H}}}[i] + \mathcal{T}_{\mathrm{S}}{}^{\gamma\lambda_i}\left(r^{(t)}[i] - y_{\mathcal{T}_{\mathrm{H}}}[i]\right)\right)_{i\in I} - r^{(t)}$.

(8) Update $x^{(t+1)}$ using (33): $\quad x^{(t+1)} = \left(1 - \dfrac{\mu_t}{2}\right) x^{(t)} + \dfrac{\mu_t}{2} q^{(t)}$.

End main iteration
Output: Denoised image $\hat{S} = \exp\left(\widetilde{W} x^{(N_{\mathrm{DR}})}\right)(1 + \psi_1(K)/2)$.

Remark 4 (Computation load) The bulk of computation of our denoising algorithm is invested in applying W and its pseudo-inverse $\widetilde{W}$. These operators are of course never constructed explicitly, rather they are implemented as fast implicit analysis and synthesis operators. Each application of W or $\widetilde{W}$ cost $\mathcal{O}(N \log N)$ for the second generation curvelet transform of an N-pixel image [15]. If we define N_{DR} and N_{FB} as the number of iterations in the Douglas-Rachford algorithm and the forward-backward sub-iteration, the computational complexity of the denoising algorithm is of order $N_{\mathrm{DR}} N_{\mathrm{FB}} 2N \log N$ operations.

6 Experiments

In all experiments carried out in this paper, our algorithm was run using second-generation curvelet tight frame along with the following set of parameters: $\forall t, \mu_t \equiv 1$, $\beta_t = 0.24$, $\gamma = 10$ and $N_{\mathrm{DR}} = 50$. The initial threshold T was set to $2\sqrt{\psi_1(K)}$. For comparison purposes, some very recent multiplicative noise removal algorithms from the literature are considered: the AA algorithm [10] minimizing the criterion in (14), and the Stein-Block denoising method [21] in the curvelet domain, applied on the log transformed image. The latter is a sophisticated shrinkage-based denoiser that thresholds the coefficients by blocks rather than individually, and has been shown to be nearly minimax over a large class of images in presence of additive bounded noise (not necessarily Gaussian nor independent). We also tried the "naive" method, called L2-TV, where $\hat{u}$ minimizes (15) and the denoised image is given after bias correcion according to (52). No without surprise, one realizes that the results are quite good, even though some persistent outliers remain quite visible. This again raises the persistent question of relevance of PSNR (or even MAE) as a measure of perceptual restoration quality. For fair comparison, the hyperparameters for all competitors were tweaked to reach their best level of performance on each noisy realization.

The denoising algorithms were tested on three images: Shepp-Logan phantom, Lena and Boat all of size 256×256 and with gray-scale in the range $[1, 256]$. For each image, a noisy observation is generated by multiplying the original image by a realization of noise according to the model in (2)-(3) with the choice $\mu = 1$ and $K = 10$. For a N-pixel noise-free image S_0 and its denoised version by any algorithm $\hat{S}$, the denoising performance is measured in terms of peak signal-to-noise ratio (PSNR) in decibels (dB)

$$\mathrm{PSNR} = 20 \log_{10} \frac{\sqrt{N}\|S_0\|_\infty}{\left\|\hat{S} - S_0\right\|} \ \mathrm{dB} \ ,$$

and mean absolute-deviation MAE

$$\mathrm{MAE} = \left\|\hat{S} - S_0\right\|_1 / N \ .$$

The results are depicted in Fig. 3, Fig. 4 and Fig. 5. Our denoiser clearly outperforms its competitors both visually and quantitatively as revealed by the PSNR and MAE values. The PSNR improvement brought by our approach is up to 4dB on the Shepp-Logan phantom, and is $\sim$ 1dB for Lena and Boat. Note also that a systematic behavior of AA algorithm is its tendency to lose some important details and the persistence of a low-frequency ghost as it can be seen on the error maps on the third row in Figs. 4 and 5.

7 Conclusions

This work proposes quite an original, efficient and fast method for multiplicative noise removal. The latter is a difficult problem that arises in various applications relevant to active imaging system, such as laser imaging, ultrasound imaging, SAR and many others. Multiplicative noise contamination involves inherent difficulties that severely restrict the main restoration algorithms.

The main ingredients of our method are: (1) consider the log-data to restore a log-image; (2) preprocess the log-fata using and under-optimal hard-thresholding of its tight frame coefficients; (3) restore the log-image using a hybrid criterion composed of an ℓ^1 data-fitting for the coefficients and a TV regularization in the log-image domain; (4) restore the sought-after image using an exponential transform along with a pertinent bias correction. The resultant algorithm is fast, its consistency and convergence are proved theoretically.

The obtained numerical results are really encouraging since they outperform the most recent methods in this field.

References

[1] http://www.cmap.polytechnique.fr/∼antonin/software/. Technical report.

[2] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions*. Dover Publications, New York, 1972.

[3] R. Acar and C. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *IEEE Transactions on Image Processing*, 10(6):1217–1229, Dec. 1994.

[4] A. Achim, A. Bezerianos, and P. Tsakalides. Novel bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Trans. Med. Imaging*, 20(8):772–783, Aug. 2001.

[5] A. Achim, E. Kuruoglu, and J. Zerubia. Sar image filtering based on the heavy-tailed rayleigh model. *IEEE Transactions on Image Processing*, 15(9):2686–2693, Sep. 2006.

[6] A. Achim, P. Tsakalides, and A. Bezerianos. Sar image denoising via bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. Remote Sens.*, 41(8):1773–1784, Aug. 2003.

[7] A. Antoniadis and Jianqing Fan. Regularization of wavelet approximations. *Journal of Acoustical Society America*, 96(455):939–967, Sep. 2001.

[8] A. Antoniadis, D. Leporini, and J.-C. Pesquet. Wavelet thresholding for some classes of non-gaussian noise. *Statistica Neerlandica*, 56(4):434–453, Dec. 2002.

[9] G. Aubert and J.-F. Aujol. Modeling very oscillating signals. Application to image processing. *Applied Mathematics and Optimization*, 51(2):163–182, Mar. /Apr. 2005.

[10] G. Aubert and J.-F. Aujol. A variational approach to remove multiplicative noise. *SIAM Journal on Applied Mathematics*, 68(4):925–946, Jan. 2008.

[11] Gilles Aubert and Pierre Kornprobst. *Mathematical problems in image processing*. Springer-Verlag, Berlin, 2 edition, 2006.

[12] J.-F. Aujol. Some first-order algorithms for total variation based image restoration. *Report CLMA N? 2008-05*, 2008.

[13] Murat Belge, Misha Kilmer, and Eric Miller. Wavelet domain image restoration with adaptive edge-preserving regularization. *IEEE Transactions on Image Processing*, 9(4):597–608, Apr. 2000.

[14] Y. Bobichon and A. Bijaoui. Regularized multiresolution methods for astronomical image enhancement. *Exper. Astron.*, (7):239–255, 1997.

[15] E. J. Candès, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Model. Simul.*, 5(3):861–899, Jan. . 2005.

[16] E. J. Candès and F. Guo. New multiscale transforms, minimum total variation synthesis. Applications to edge-preserving image reconstruction. *Signal Processing*, 82, Mar. 2002.

[17] A. Chambolle and J. Darbon. On total variation minimization and surface evolution using parametric maximum flows. Technical report, CMAP-08, Ecole Polytechnique, France, 2008.

[18] Antonin Chambolle. An algorithm for total variation minimization and application. *Journal of Mathematical Imaging and Vision*, 20(1), Jan. -Mar. 2004.

[19] T.F. Chan and H.M. Zhou. Total variation improved wavelet thresholding in image compression. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 391–394. IEEE, 2000.

[20] Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, Feb. 1997.

[21] C. Chesneau, J. Fadili, and J.-L. Starck. Stein block thresholding for image denoising. Technical report.

[22] R. R. Coifman and D. Donoho. Translation-invariant de-noising. Technical Report Report 475, Stanford University, Dept. of Statistics, 1995.

[23] R. R. Coifman and A. Sowa. Combining the calculus of variations and wavelets for image enhancement. *Applied and Computational Harmonic Analysis*, 9, 2000.

[24] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5), Dec. 2004.

[25] P. L. Combettes and J.-C. Pesquet. A Douglas-Rachford splittting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, 2007.

[26] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Multiscale Model. Simul.*, 4(4):1168–1200, 2005.

[27] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[28] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of Acoustical Society America*, 90, 1995.

[29] S. Durand and Nikolova M. Denoising of frame coefficients using l1 data-fidelity term and edge-preserving regularization. *SIAM Journal on Multiscale Modeling and Simulation*, 6(2):547–576, 2007.

[30] Sylvain Durand and Jacques Froment. Reconstruction of wavelet coefficients using total variation minimization. *SIAM Journal on Scientific Computing*, 24(5):1754–1767, 2003.

[31] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming: Series A and B*, 55(3):293–318, July . 1992.

[32] J. Eckstein and B. F. Svaiter. A family of projective splitting methods for the sum of two maximal monotone operators. *Math. Program., Ser. B*, 111(1), Jan. . 2008.

[33] J. Froment and S. Durand. Artifact free signal denoising with wavelets. In *Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 6, 2001.

[34] S. Fukuda and H. Hirosawa. Suppression of speckle in synthetic aperture radar images using wavelet. *Int. J. Remote Sens.*, 19(3):507–519, 1998.

[35] D. Gabay. *Applications of the method of multipliers to variational inequalities*. M. Fortin and R. Glowinski, editors, North-Holland, Amsterdam, 1983.

[36] Yu-Mei Huang, Michael K. Ng, and You-Wei Wen. A new total variation method for multiplicative noise removal. Technical report, Hong Kong Baptist University, http://www.math.hkbu.edu.hk/ICM/pdf/08-06.pdf, 2008.

[37] K. Krissian, C.-F. Westin, R. Kikinis, and K. G. Vosburgh. Oriented speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 16(5):1412–1424, May 2007.

[38] P.-L. Lions. Une méthode itérative de resolution d'une inéquation variationnelle. *Israel Journal of Mathematics*, 31(2):204–208, June . 1978.

[39] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, Dec. . 1979.

[40] F. Malgouyres. Mathematical analysis of a model which combines total variation and wavelet for image restoration. *Journal of information processes*, 2(1):1–10, 2002.

[41] François Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE Transactions on Image Processing*, 11(12):1450–1456, Dec. 2002.

[42] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *CRAS Sér. A Math.*

[43] P. Moulin and J. Liu. Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors. *IEEE Transactions on Image Processing*, 45(3):909–919, Apr. 1999.

[44] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM Journal on Numerical Analysis*, 40(3):965–994, 2002.

[45] M. Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20(1):99-120, Jan. -Mar. 2004.

[46] M. Nikolova. Weakly constrained minimization. Application to the estimation of images and signals involving constant regions *Journal of Mathematical Imaging and Vision*, 21(2):155-175, Sep. 2004.

[47] M. Nikolova. Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *SIAM Journal on Multiscale Modeling and Simulation*, 4(3):960–991, 2005.

[48] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 72, 1979.

[49] A. Pizurica, A. M. Wink, E. Vansteenkiste, W. Philips, and J.B.T.M. Roerdink. A review of wavelet denoising in mri and ultrasound brain imaging. *Current Medical Imaging Reviews*, 2(2):247–260, 2006.

[50] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery. Numerical recipes, the art of scientific computing. *Cambridge Univ. Press, New York*, 1992.

[51] R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, Aug. 1976.

[52] L. Rudin, P.-L. Lions, and S. Osher. *Multiplicative denoising and deblurring: Theory and algorithms.*, pages 103–119. Springer, Editors: S. Osher and N. Paragios, 2003.

[53] L. Rudin, S. Osher, and C. Fatemi. Nonlinear total variation based noise removal algorithm. *Physica*, 60 D:259–268, 1992.

[54] E. P. Simoncelli. *Bayesian denoising of visual images in the wavelet domain*. Lecture Notes in Statistics, Vol. 41. Springer Verlag: Berlin, 1999.

[55] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coding. In *Proceedings of the IEEE International Conference on Image Processing*, pages 379–382, Lausanne, Switzerland, Sep. 1996.

[56] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Winston, Washington DC, 1977.

[57] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control and Optimization*, 29(1):119–138, 1991.

[58] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(1):431–446, 2000.

[59] M. Tur, C. Chin, and J.W. Goodman. When is speckle noise multiplicative? *Applied Optics*, 21(7):11571159, April 1982.

[60] F. Ulaby and M. C. Dobson. *Handbook of Radar Scattering Statistics for Terrain*. Norwood, MA: Artech House, 1989.

[61] C. R. Vogel and M. E. Oman. Iterative method for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.

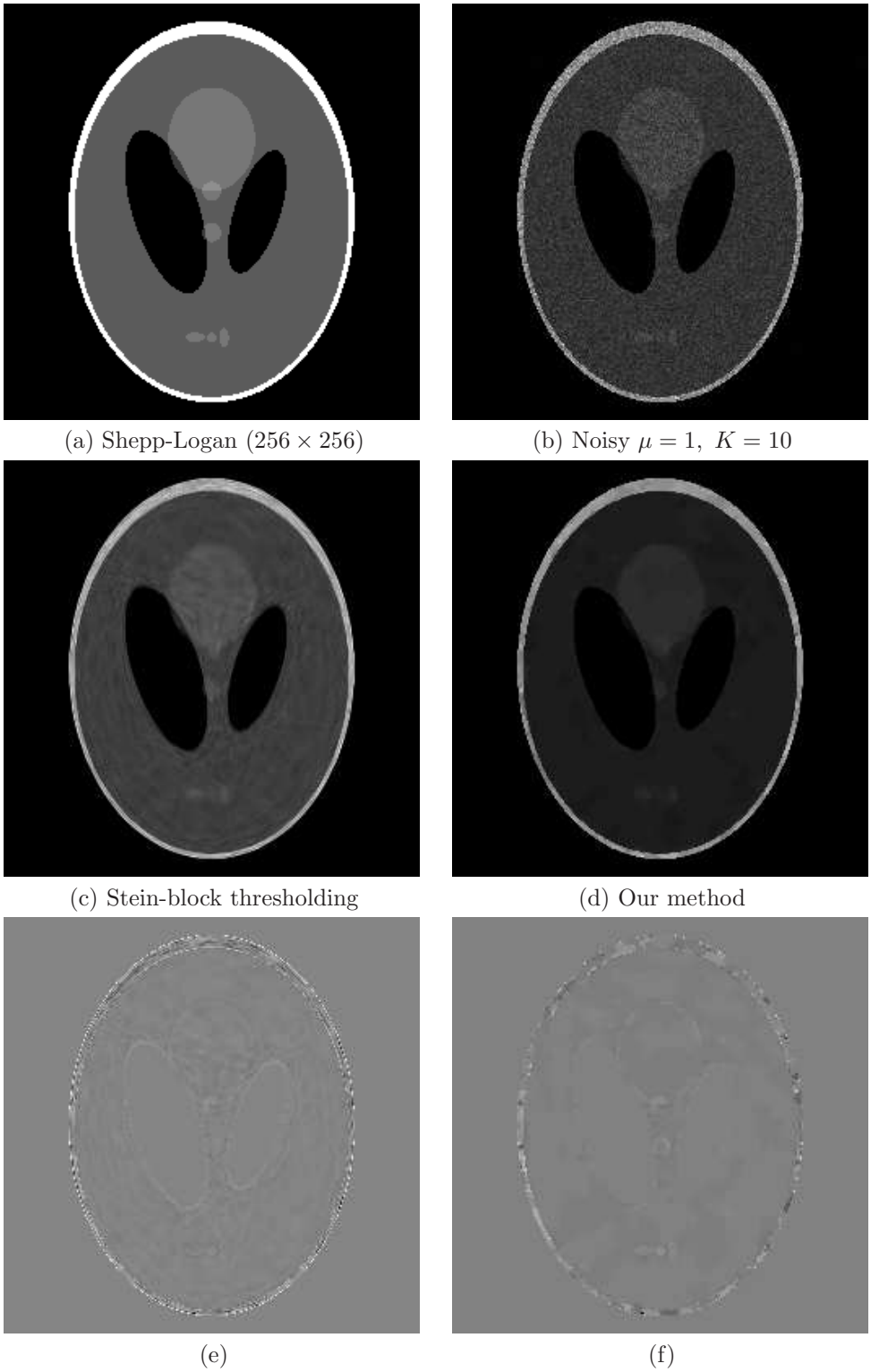
[62] M. Walessa and M. Datcu. Model-based despeckling and information extraction from sar images. *IEEE Trans. Geosci. Remote Sens.*, 38(9):2258–2269, Sep. 2000.

[63] G. Wang, J. Zhang, and G.-W. Pan. Solution of inverse problems in image processing by wavelet expansion. *IEEE Transactions on Image Processing*, 4(5):579–593, May 1995.

[64] H. Xie, L. E. Pierce, and F. T. Ulaby. SAR speckle reduction using wavelet denoising and markov random field modeling. *IEEE Trans. Geosci. Remote Sensing*, 40(10):2196–2212, Oct. 2002.

[65] Y. Yu and S. T. Acton. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 11(11):1260–1270, Nov. 2002.

[66] M. Zhu, S. J. Wright and T. F. Chan Duality-Based Algorithms for Total-Variation-Regularized Image Restoration. Technical report, CAM 08-33, UCLA, October, 2008.

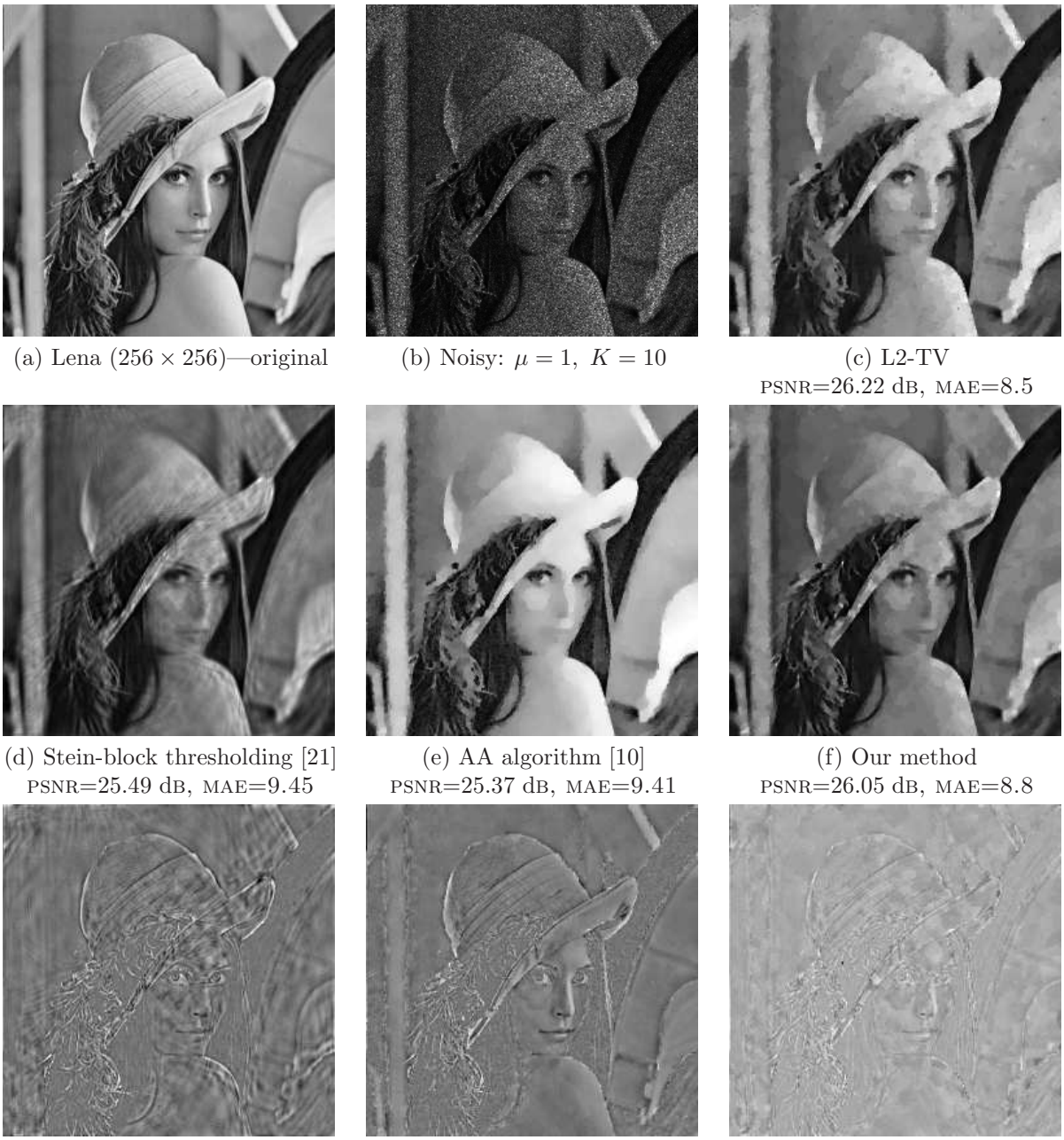


(a) Shepp-Logan (256×256) (b) Noisy $\mu = 1,\ K = 10$

(c) Stein-block thresholding (d) Our method

(e) (f)

Figure 3: Performance comparison with Shepp-Logan phantom (256×256). (a) Original. (b) Noisy $\mu = 1, K = 10$. (c) Denoised with Stein-block thresholding in the curvelet domain [21] PSNR=24.73dB, MAE=4. (d) Denoised with our algorithm PSNR=31.25dB, MAE=1.87. (e)-(f) Errors (restored − original) for (c)-(d).



(a) Lena (256×256)—original

(b) Noisy: $\mu = 1,\ K = 10$

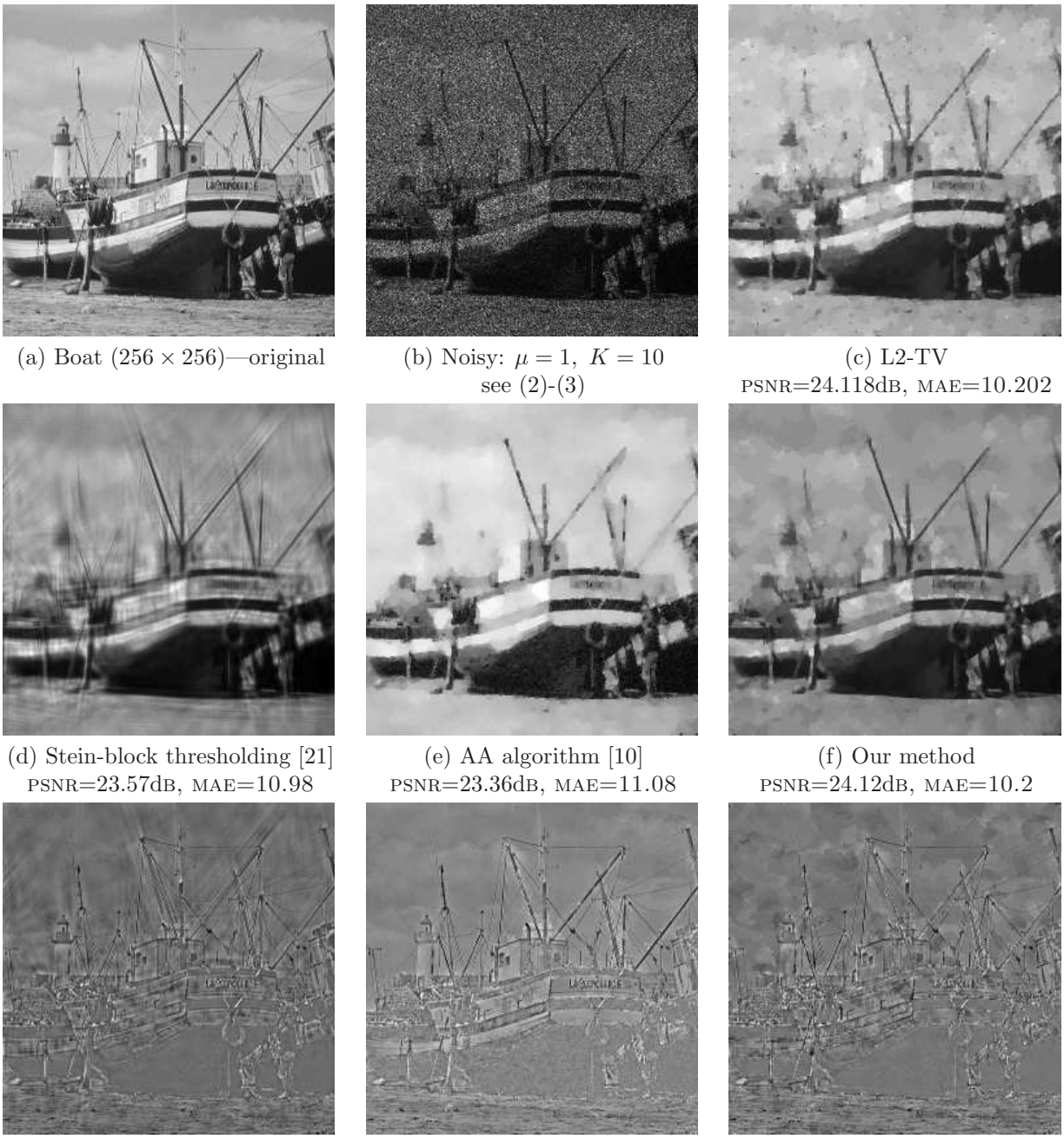
(c) L2-TV
PSNR=26.22 dB, MAE=8.5

(d) Stein-block thresholding [21]
PSNR=25.49 dB, MAE=9.45

(e) AA algorithm [10]
PSNR=25.37 dB, MAE=9.41

(f) Our method
PSNR=26.05 dB, MAE=8.8

Figure 4: Comparative restoration of the noisy Lena in (b) using modern methods. Note that the algorithm in (c) is initialized with the log-data and that the restoration in (d) is done in the curvelet domain. The images on the last row show the error (restored − original) for (d), (e) and (f).



(a) Boat (256×256)—original

(b) Noisy: $\mu = 1,\ K = 10$ see (2)-(3)

(c) L2-TV PSNR=24.118dB, MAE=10.202

(d) Stein-block thresholding [21] PSNR=23.57dB, MAE=10.98

(e) AA algorithm [10] PSNR=23.36dB, MAE=11.08

(f) Our method PSNR=24.12dB, MAE=10.2

Figure 5: Restoration of (b) using contemporary methods. The last row shows the error images, namely (restored − original) for (d), (e) and (f).