



Multiplicative Noise Cleaning via a Variational Method Involving Curvelet Coefficients

Sylvain Durand¹, Jalal Fadili², and Mila Nikolova³

¹ M.A.P. 5 - CNRS, University Paris Descartes, France
Sylvain.Durand@mi.parisdescartes.fr

<http://www.math-info.univ-paris5.fr/~s>

² GREYC CNRS-ENSICAEN-Université de Caen, France
Jalal.Fadili@greyc.ensicaen.fr

<http://www.greyc.ensicaen.fr/~jfadili/>

³ CMLA - CNRS, ENS Cachan, PRES UniverSud, France
Mila.Nikolova@cmla.ens-cachan.fr

<http://www.cmla.ens-cachan.fr/~nikolova/>

Abstract. Classical ways to denoise images contaminated with multiplicative noise (e.g. speckle noise) are filtering, statistical (Bayesian) methods, variational methods and methods that convert the multiplicative noise into additive noise (using a logarithmic function) in order to apply a shrinkage estimation for the log-image data and transform back the result using an exponential function.

We propose a new method that involves several stages: we apply a reasonable under-optimal hard-thresholding on the curvelet transform of the log-image; the latter is restored using a specialized hybrid variational method combining an ℓ^1 data-fitting to the thresholded coefficients and a Total Variation regularization (TV) in the image domain; the restored image is an exponential of the obtained minimizer, weighted so that the mean of the original image is preserved. The minimization stage is realized using a properly adapted fast Douglas-Rachford splitting. The existence of a minimizer of our specialized criterion and the convergence of the minimization scheme are proved. The obtained numerical results outperform the main alternative methods.

1 Introduction

In many active imaging systems (e.g. synthetic aperture radar, laser or ultrasound imaging), the data for the unknown image $S_0 : \Omega \rightarrow \mathbb{R}_+$, $\Omega \subset \mathbb{R}^2$, are severely corrupted with multiplicative noise. Then several independent measurements for the same image are needed:

$$S_k = S_0 \eta_k + n_k, \quad \forall k \in \{1, \dots, K\}, \quad (1)$$

where $\eta_k : \Omega \rightarrow \mathbb{R}_+$, and n_k represent the multiplicative and a typically zero-mean additive noise, $\forall k$. Commonly (see e.g. [27]) η_k is modeled as a one-sided exponential probability density function (pdf) (*cf.* Fig. 1(a)): $\text{pdf}(\eta_k) =$

$\mu e^{-\mu\eta_k} \mathbb{1}_{\mathbb{R}_+}(\eta_k)$ for $\mu = 1$. In practice, one takes an average of all measurements, see e.g. Fig. 2(b). Since $\frac{1}{K} \sum_{k=1}^K n_k \approx 0$, the data read (*cf.* e.g. [27, 1, 30]):

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

Usually all η_k are independent. Denoting by Γ the usual Gamma-function, the mean of the noise η in (2) has a Gamma distribution (*cf.* Fig. 1(b)):

$$\eta = \frac{1}{K} \sum_{k=1}^K \eta_k : \quad \text{pdf}(\eta) = \frac{K^K \eta^{K-1} \exp(-K\eta)}{\Gamma(K)} . \tag{3}$$

Various adaptive filters have been proposed, see e.g. [31, 17]: they work well when the noise is moderate or weak, i.e. for K large. Bayesian, variational or diffusion-based methods have been proposed as well; see e.g. [28, 24, 18, 2]. Numerous methods convert the multiplicative noise into additive noise by

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

see e.g. [16, 30, 1, 23]. Then the pdf of n reads (*cf.* Fig. 1(c)):

$$n = \log \eta : \quad \text{pdf}(n) = K^K (\Gamma(K))^{-1} \exp(-K(n - e^n)) . \tag{5}$$

One can prove that $\mathbb{E}[n] = \psi_0(K) - \log K$ and $\text{Var}[n] = \psi_1(K)$, where $\psi_k(z) = \left(\frac{d}{dz}\right)^{k+1} \log \Gamma(z)$ is the polygamma function. A common strategy is to decompose the log-data v into a multiscale frame for $L^2(\mathbb{R}^2)$ (an over-complete basis), say $W \equiv \{w_i : i \in I\}$ where I is a set of indexes:

$$y = Wv = Wu_0 + Wn. \tag{6}$$

By the Central Limit Theorem, the noise Wn in y is nearly Gaussian — *cf.* Fig. 1(d). Then coefficients y are denoised using shrinkage estimators $\mathcal{T} : \mathbb{R} \rightarrow \mathbb{R}$:

$$y_{\mathcal{T}}[i] = \mathcal{T}((Wv)[i]), \quad \forall i \in I. \tag{7}$$

Shrinkage functions designed for multiplicative noise were proposed e.g. in [30, 1]. Let $\widetilde{W} \equiv \{\widetilde{w}_i : i \in I\}$ be a left inverse of W . Then a denoised log-image $v_{\mathcal{T}}$ reads

$$v_{\mathcal{T}} = \sum_{i \in I} \mathcal{T}((Wv)[i]) \widetilde{w}_i = \sum_{i \in I} \mathcal{T}(y[i]) \widetilde{w}_i. \tag{8}$$

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

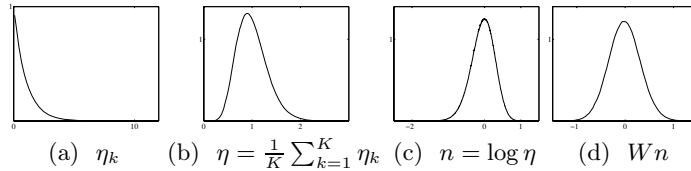


Fig. 1. Noise distributions

Our approach. We apply (4) and consider a tight-frame transform of the log-data. The restored log-image (section 2) minimizes 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. The minimization (section 3) uses a specialized Douglas-Rachford splitting. The full algorithm, involving a bias correction, is given in section 4. Experiments are presented in section 5.

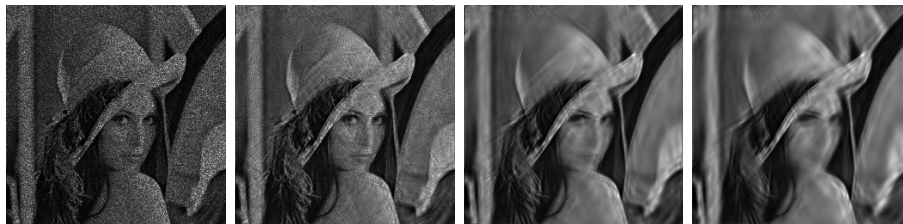
Some notations. (\cdot^T) means transposed, (\cdot^*) means convex conjugate and (\cdot^\dagger) means adjoint.

2 Restoration of the Log-Image

Here we consider how to restore a good log-image given data $v : \Omega_{\text{mega}} \rightarrow \mathbb{R}$ obtained using (4). We focus on methods which, for a given preprocessed data set, lead to convex optimization problems. We comment only variational methods and shrinkage estimators since they underly our specialized hybrid objective function.

2.1 Drawbacks of Shrinkage Restoration and Variational Methods

Shrinkage restoration. The main problems with these methods, sketched in (7)-(8), is that shrinking large coefficients entails an erosion of the spiky features, while shrinking small coefficients yields Gibbs-like oscillations in the vicinity of edges and a loss of details in the textured area. On the other hand, if shrinkage is insufficient, some coefficients bearing mainly noise can remain almost unchanged—we call such coefficients *outliers*—and (8) shows that they yield artifacts with the shape of the functions \tilde{w}_i , see Fig. 2. Even though various improvements were brought, these artifacts remain visible—see the results on Fig. 3(d) and Fig. 4(c) in Section 5 using the very recent Stein-block thresholding [8].



(a) Noisy, $K = 10$ (b) $T = 2\sqrt{\text{Var}[n]}$ (c) $T = 4\sqrt{\text{Var}[n]}$ (d) $T = 6\sqrt{\text{Var}[n]}$

Fig. 2. (a) Noisy Lena obtained according to (1)-(2) for $K = 10$. (b)-(d) Restorations $\exp v_{\mathcal{T}_H}$ where data v are denoised by hard-thresholding of its curvelet coefficients, see (12)-(13), for different choices of T .

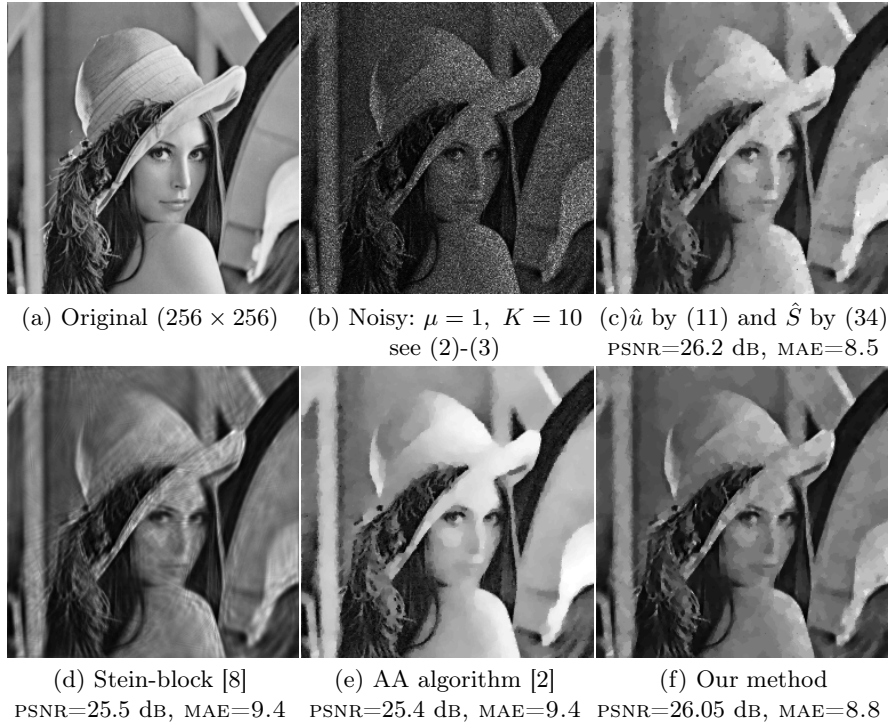


Fig. 3. Restoration of (b) using modern methods. Note that (c) is a slightly improved version of [26] and that the restoration in (d) is done in the curvelet domain.

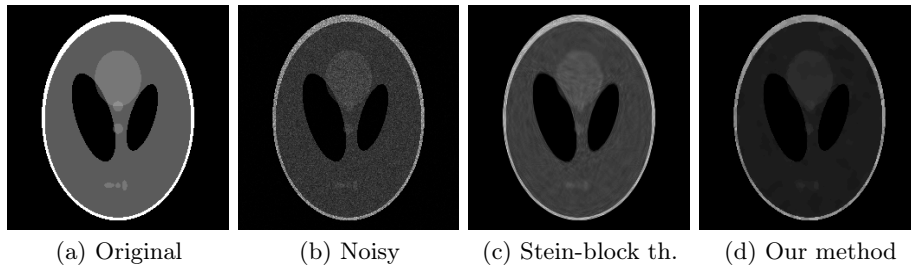


Fig. 4. (a) Shepp-Logan phantom (256×256). (b) Noisy, $K = 10$. (c) Denoised with Stein-block thresholding in the curvelet domain [8] PSNR=24.73dB, MAE=4. (d) Denoised with our algorithm PSNR=31.25dB, MAE=1.87.

Variational methods. In these methods, the restored function minimizes a criterion \mathcal{F}_v of the form

$$\mathcal{F}_v(u) = \rho \int_{\Omega} \psi(u(t), v(t)) dt + \int_{\Omega} \varphi(|\nabla u(t)|) dt, \quad (9)$$

where $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ measures closeness to data and $\varphi(|\nabla u(\cdot)|)$ introduces priors via a trade-off parameter $\rho > 0$. A classical choice is $\psi = (u(\cdot) - v(\cdot))^2$. It is usually required that the potential function $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ promotes images involving edges. Analysing the minimizers of \mathcal{F}_v as solutions of PDE's on Ω , Rudin, Osher and Fatemi [25] exhibited that $\varphi(|\nabla u(t)|) = |\nabla u(t)|$, leads to such images, where for any $z(t) = (z_1(t), z_2(t)) \in \mathbb{R}^2, t \in \Omega$, one sets $|z(t)| \stackrel{\text{def}}{=} \sqrt{z_1(t)^2 + z_2(t)^2}$. The resulting regularization term is known as Total Variation (TV) and will be denoted by $\|\cdot\|_{\text{TV}}$. However, whatever smooth data-fitting is chosen, this regularization yields images containing numerous constant regions (called stair-casing effect), hence textures and fine details are removed, see [22]. The method in [2] is of this kind and operates in the image domain; the fitting term is derived from (3) and the denoised image \hat{S} , defined by

$$\hat{S} = \arg \min_{\Sigma} \mathcal{F}_S \text{ for } \mathcal{F}_S(\Sigma) = \rho(K) \int (\log \Sigma(t) + S(t)/\Sigma(t)) dt + \|\Sigma\|_{\text{TV}}, \quad (10)$$

exhibits constant regions (see section 5). In [26], the regularization $\|\Sigma\|_{\text{TV}}$ is changed into $\|\log \Sigma\|_{\text{TV}}$ so as to reformulate the model as a convex problem, and not to over smooth the image parts with higher gray values. To recover the denoised image, we applied $\hat{S} \propto \exp(\hat{u})$ for

$$\hat{u} = \arg \min_u \text{ where } F_v(u) = \rho \|u - v\|^2 + \|u\|_{\text{TV}}. \quad (11)$$

Following [25], various edge-preserving convex functions φ have been proposed; see [3] for a recent overview. Even though $\varphi'(0) = 0$ alleviates stair-casing, a systematic drawback of the resulting restored images is that the amplitude of edges is underestimated; thus neat edges or spiky areas are subjected to erosion.

2.2 Hybrid Methods

Hybrid methods, see e.g. [9, 19, 5, 14], combine the information contained in the large coefficients $y[i]$ obtained according to (6) with priors directly on the image u . They amount to define the restored function \hat{u} by

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

Using an edge-preserving regularization, such as $\Phi = \text{TV}$ is a pertinent choice. 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, choosing 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 good compromise that we adopt is to determine $(\mu_i)_{i \in I}$ based both on the data and on the prior term.

2.3 A Specialized Hybrid Criterion

Given the log-data v obtained by (4), we apply a frame transform as in (6) to get $y = Wv = Wu_0 + Wn$. The noise contained in the i -th datum reads $\langle n, w_i \rangle$.

The low frequency approximation coefficients carry important information on the image. Therefore, a good choice is to keep them intact at this stage. Let $I_* \subset I$ denote the subset of all such elements of the frame. Then we apply a hard-thresholding operator \mathcal{T}_H [12] to all coefficients $I \setminus I_*$:

$$y_{\mathcal{T}_H}[i] \stackrel{\text{def}}{=} \mathcal{T}_H(y[i]), \quad \forall i \in I \setminus I_*, \quad \text{where} \quad \mathcal{T}_H(t) = \begin{cases} 0 & \text{if } |t| \leq T, \\ t & \text{otherwise,} \end{cases} \quad (12)$$

where T is an *underoptimal* threshold in order to preserve the information relevant to edges and to some fine details in textured areas, contained in the small coefficients. Let us consider

$$v_{\mathcal{T}_H} = \sum_{i \in I_1} Wv[i] \tilde{w}_i, \quad \text{where} \quad I_1 = \{i \in I : |y[i]| > T\} \cup I_*. \quad (13)$$

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

The denoised coefficients, denoted by \hat{x} , are obtained based on the under-thresholded data $y_{\mathcal{T}_H}$. We focus on hybrid methods of the form: $\hat{x} = \arg \min_x F(x)$ for $F(x) = \Psi(x, y_{\mathcal{T}_H}) + \Phi(\tilde{W}x)$, where Ψ is a data-fitting term in the frame domain and Φ is an edge-preserving regularization term in the log-image domain. Let us denote

$$I_0 = I \setminus I_1 = \{i \in I \setminus I_* : |y[i]| \leq T\}. \quad (14)$$

Coefficients $y[i]$ for $i \in I_0$ can be of the two types. 1. Coefficients $y[i]$ bearing mainly noise—then the best choice is $\hat{x}[i] = 0$; 2. Coefficients $y[i]$ relevant 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 Φ . Note that a careful restoration must find a nonzero $\hat{x}[i]$ in order to avoid Gibbs-like oscillations in \hat{u} .

Coefficients $y[i]$ for $i \in I_1$ are of the following two types. 1. 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. 2. Coefficients highly contaminated by noise, i.e. $|y[i]| \gg |\langle w_i, u_0 \rangle|$. We call them *outliers* because if we had $\hat{x}[i] = y[i]$, then \hat{u} would contain an artifact with the shape of \tilde{w}_i since by (13) we get $v_{\mathcal{T}_H} = \sum_{j \in I} \hat{x}[j] \tilde{w}_j + y[i] \tilde{w}_i$. 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 is expected to achieve: restored coefficients $\hat{x}[i]$ have to fit $y_{\mathcal{T}_H}[i]$ exactly if they are coherent with the prior Φ , otherwise they have to be restored according to Φ . Since [21] it is known that such requirements can be satisfied by criteria F where Ψ is non-smooth at the origin (e.g. ℓ^1), see also [13]. For these reasons, we focus on

$$F(x) = \Psi(x) + \Phi(x), \quad (15)$$

where, for $\Lambda = \text{diag}(\lambda_i)_{i \in I}$,

$$\Psi(x) = \sum_{i \in I_1 \cup I_*} \lambda_i |(x - y)[i]| + \sum_{i \in I_0} \lambda_i |x[i]| = \|\Lambda(x - y_{\mathcal{T}_H})\|_1, \quad (16)$$

$$\Phi(x) = \int_{\Omega} |\nabla \widetilde{W}x| \, ds = \|\widetilde{W}x\|_{\text{TV}}. \quad (17)$$

In the pre-processing step (12) we do not recommend the use of a shrinkage function other than \mathcal{T}_H since it will alter *all* the data coefficients without restoring them faithfully. Via \mathcal{T}_H , we base our restoration on data $y_{\mathcal{T}_H}$ where all non-thresholded coefficients keep the original information on the sought-after image.

The theorem stated next addresses the existence and the uniqueness of a minimizer for F . Given y , let G_y be the (convex) set of all minimizers of F :

$$G_y \stackrel{\text{def}}{=} \{\hat{x} \in \ell^2(I) : F(\hat{x}) = \min_{x \in \ell^2(I)} F(x)\}. \quad (18)$$

Theorem 1. [13] *For $y \in \ell^2(I)$ and $T > 0$ given, consider F as defined in (15), where $\Omega \in \mathbb{R}^2$ is open, bounded and its boundary $\partial\Omega$ is Lipschitz. Suppose that $\{w_i\}_{i \in I}$ is a frame of $L^2(\Omega)$ and the operator \widetilde{W} is the pseudo-inverse of W . Assume also that $\lambda_{\min} = \min_{i \in I} \lambda_i > 0$. Then G_y is nonempty, and for all $\hat{x}_1, \hat{x}_2 \in G_y$, $\nabla \widetilde{W}\hat{x}_1 \propto \nabla \widetilde{W}\hat{x}_2$, a.e. on Ω .*

In words, $\hat{S}_1 = \widetilde{W}\hat{x}_1$ and $\hat{S}_2 = \widetilde{W}\hat{x}_2$ have the same level lines, i.e. they differ by a local change of contrast; the latter is usually invisible to the naked eye.

The choice of λ_i is investigated in [13]. Following this analysis, we use only two values for λ_i , depending only on the set I_ϵ the index i belongs to. We focus on 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 $\Gamma_0(\mathcal{H})$ denote the class of proper lower-semicontinuous convex functions on a Hilbert space \mathcal{H} . Now we focus on the minimization problem

$$\text{find } \hat{x} \text{ such that } F(\hat{x}) = \min_x F \text{ for } F = \Psi + \Phi, \quad (19)$$

where Ψ and Φ are defined in (16)-(17). Clearly, $\Psi, \Phi \in \Gamma_0(\ell_2(I))$, hence $F \in \Gamma_0(\ell_2(I))$. The set G_y in (18) is non-empty by Theorem 1 and can be rewritten as $G_y = \{x \in \ell^2(I) | x \in (\partial F)^{-1}(0)\}$, where ∂F stands for subdifferential. Minimizing F amounts to finding a solution to the fixed point equation

$$x = (\text{Id} + \gamma \partial F)^{-1}(x), \quad (20)$$

where $(\text{Id} + \gamma \partial F)^{-1}$ is the *resolvent operator* associated to ∂F , $\gamma > 0$ is the proximal stepsize and Id is the identity map on $\ell^2(I)$. Since $(\text{Id} + \gamma(\partial\Psi + \partial\Phi))^{-1}$ cannot be calculated in closed-form, we focus on splitting methods that use separately the resolvent operators $(\text{Id} + \gamma \partial\Psi)^{-1}$ and $(\text{Id} + \gamma \partial\Phi)^{-1}$.

3.1 Specialized Douglas-Rachford (D-R) Splitting Algorithm

The D-R family is the most general class of monotone operator splitting methods. Given a sequence $\mu_t \in (0, 2)$, D-R methods can be expressed via the recursion

$$x^{(t+1)} = \left[\left(1 - \frac{\mu_t}{2}\right) \text{Id} + \frac{\mu_t}{2} (2(\text{Id} + \gamma \partial \Psi)^{-1} - \text{Id}) \circ (2(\text{Id} + \gamma \partial \Phi)^{-1} - \text{Id}) \right] x^{(t)}. \tag{21}$$

Since problem (19) has solutions, we have the following convergence result:

Theorem 2. *Let $\gamma > 0$ and $\mu_t \in (0, 2)$ be such that $\sum_{t \in \mathbb{N}} \mu_t(2 - \mu_t) = +\infty$. Take $x^{(0)} \in \ell^2(I)$ and consider the sequence of iterates defined by (21). Then, $(x^{(t)})_{t \in \mathbb{N}}$ converges weakly to some point $\hat{x} \in \ell^2(I)$ and $(\text{Id} + \gamma \partial \Phi)^{-1}(\hat{x}) \in G_y$.*

The statement follows from [10, Corollary 5.2]. The sequence $\mu_t = 1, \forall t \in \mathbb{N}$ fits.

3.2 Proximal Calculus

Proximity operators, invented in [20], generalize convex projection.

Definition 1 (Moreau [20]). *Let $\varphi \in \Gamma_0(\mathcal{H})$. Then $\forall x \in \mathcal{H}$ the function $z \mapsto \varphi(z) + \|x - z\|^2 / 2$, for $z \in \mathcal{H}$, achieves its infimum at a unique point denoted by $\text{prox}_\varphi x$. The relevant operator $\text{prox}_\varphi : \mathcal{H} \rightarrow \mathcal{H}$ is the proximity operator of φ .*

By the minimality condition for prox_φ , it is easy to see that $\forall x, p \in \mathcal{H}$ we have $p = \text{prox}_\varphi x \iff x - p \in \partial \varphi(p) \iff (\text{Id} + \partial \varphi)^{-1} = \text{prox}_\varphi$. By introducing the reflection operator $\text{rprox}_\varphi \stackrel{\text{def}}{=} 2\text{prox}_\varphi - \text{Id}$, the D-R iteration (21) reads

$$x^{(t+1)} = \left[\left(1 - \frac{\mu_t}{2}\right) \text{Id} + \frac{\mu_t}{2} \text{rprox}_{\gamma \Psi} \circ \text{rprox}_{\gamma \Phi} \right] x^{(t)}. \tag{22}$$

Proximity operator of Ψ

Lemma 1. *Let $x \in \ell^2(I)$. Then $\text{prox}_{\gamma \Psi}(x) = \left(y_{\mathcal{T}_H}[i] + \mathcal{T}_S^{\gamma \lambda_i} (x[i] - y_{\mathcal{T}_H}[i]) \right)_{i \in I}$, where $\mathcal{T}_S^{\gamma \lambda_i}(z[i]) = \max \{0, z[i] - \gamma \lambda_i \text{sign}(z[i])\}$.*

The proof is quite standard and can be found in our Report [15]. Note that

$$\text{rprox}_{\gamma \Psi}(x) = 2 \left(y_{\mathcal{T}_H}[i] + \mathcal{T}_S^{\gamma \lambda_i} (x[i] - y_{\mathcal{T}_H}[i]) \right)_{i \in I} - x. \tag{23}$$

Proximity operator of Φ . Clearly, $\Phi(x) = \|\cdot\|_{\text{TV}} \circ \widetilde{W}(x)$. Computing $\text{prox}_{\gamma \Phi}$ for an arbitrary \widetilde{W} may be intractable. We assume that

- (w1) $\widetilde{W} : \ell^2(I) \rightarrow L^2(\Omega)$ is surjective;
- (w2) $\widetilde{W}W = \text{Id}$ and $\widetilde{W} = c^{-1}W^*$ for $0 < c < \infty$; note that $W^*W = c \text{Id}$;
- (w3) \widetilde{W} is bounded.

Let $\mathcal{X} = L^2(\Omega) \times L^2(\Omega)$, $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ be the inner product in \mathcal{X} and $\|\cdot\|_p$, $p \in [1, \infty]$ the L^p -norm on \mathcal{X} . Define $\overline{B}_\infty^\gamma(\mathcal{X})$ as the γ -radius closed L^∞ -ball in \mathcal{X} , $\overline{B}_\infty^\gamma \stackrel{\text{def}}{=} \{z \in \mathcal{X} : \|z\|_\infty \leq \gamma\} = \{z = (z_1, z_2) \in \mathcal{X} : |z(t)| \leq \gamma, \forall t \in \Omega\}$, and $P_{\overline{B}_\infty^\gamma(\mathcal{X})} : \mathcal{X} \rightarrow \overline{B}_\infty^\gamma(\mathcal{X})$ the associated projector.

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

$$\text{prox}_{\gamma\Phi}(x) = \left(\text{Id} - W \circ \left(\text{Id} - \text{prox}_{c^{-1}\gamma\|\cdot\|_{\text{TV}}} \right) \circ \widetilde{W} \right) (x) ; \quad (24)$$

$$\text{prox}_{c^{-1}\gamma\|\cdot\|_{\text{TV}}}(u) = u - P_C(u) , \quad (25)$$

$$\text{where } C = \left\{ \text{div}(z) \in L^2(\Omega) \mid z \in \mathcal{C}_c^\infty(\Omega \times \Omega), z \in \overline{B}_\infty^{\gamma/c}(\mathcal{X}) \right\}. \quad (26)$$

Sketch of the proof. By (w1), $\text{range}(\widetilde{W}) = L^2(\Omega)$. Using that $\text{domain}(\|\cdot\|_{\text{TV}}) = L^2(\Omega)$, we find $\text{cone}(\text{dom}\|\cdot\|_{\text{TV}} - \text{range}\widetilde{W}) = \{0\}$. Statement (i) follows from applying [11, Proposition 11] whose requirements are satisfied.

If $\varphi \in \Gamma_0(L^2(\Omega))$ and φ^* is its convex conjugate, the Moreau decomposition [20, Proposition 4.a] asserts

$$\text{prox}_\varphi + \text{prox}_{\varphi^*} = \text{Id} . \quad (27)$$

Since the conjugate function of a norm is the indicator function of the ball of its dual norm, $(c^{-1}\gamma\|\cdot\|_{\text{TV}})^*(z) = 0$ if $z \in C$, $+\infty$ if $z \notin C$. where C is given in (26). Using Definition 1, $\text{prox}_{(c^{-1}\gamma\|\cdot\|_{\text{TV}})^*} = P_C$. Identifying $c^{-1}\gamma\|\cdot\|_{\text{TV}}$ with φ and $(c^{-1}\gamma\|\cdot\|_{\text{TV}})^*$ with φ^* , equation (27) leads to (ii)¹. \diamond

From (24)-(25) we easily find that

$$\text{rprox}_{\gamma\Phi}(x) = \left(\text{Id} - 2W \circ P_C \circ \widetilde{W} \right) (x) . \quad (28)$$

Calculation of the projection P_C in (25) on a discrete grid. In this case, W is an $M \times N$ tight frame with $M = \#I \gg N = \#\Omega$ and assumption (w2) reads $\widetilde{W}W = \text{Id}$ and $\widetilde{W} = c^{-1}W^T$, $c \in (0, \infty)$ hence $W^TW = c \text{Id}$. The discrete counterpart of \mathcal{X} is $\mathcal{X} = \ell^2(\Omega) \times \ell^2(\Omega)$. We denote the discrete gradient by $\ddot{\nabla}$ (cf. [6] or [29]) and the discrete divergence $\text{Div} : \mathcal{X} \rightarrow \ell^2(\Omega)$ is defined as $\text{Div} = -\ddot{\nabla}^*$. Moreover, C in (26) admits a simpler expression:

$$C = \left\{ \text{Div}(z) \in \ell^2(\Omega) \mid z \in \overline{B}_\infty^{\gamma/c}(\mathcal{X}) \right\} , \quad (29)$$

where $\overline{B}_\infty^{\gamma/c}(\mathcal{X})$ is defined using the new discrete notations. The projection P_C in (25) does not admit an explicit form so we provide an iterative scheme for its calculation in the next lemma.

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

$$z^{(t+1)} = P_{\overline{B}_\infty^1(\mathcal{X})} \left(z^{(t)} + \beta_t \ddot{\nabla} \left(\text{Div}(z^{(t)}) - cu/\gamma \right) \right) \quad (30)$$

$$\text{for } 0 < \inf_t \beta_t \leq \sup_t \beta_t < 1/4 \quad (31)$$

$$\text{where } \forall (i, j) \in \Omega , \quad P_{\overline{B}_\infty^1(\mathcal{X})}(z)[i, j] = \begin{cases} z[i, j] & \text{if } |z[i, j]| \leq 1; \\ z[i, j]/|z[i, j]| & \text{otherwise .} \end{cases} \quad (32)$$

¹ Note that our argument (27) to compute $\text{prox}_{c^{-1}\gamma\|\cdot\|_{\text{TV}}}(u)$ is not used in [6], which instead uses conjugates and bi-conjugates of the objective function.

Then

- (i) $(z^{(t)})_{t \in \mathbb{N}}$ converges to a point $\hat{z} \in \overline{B}_\infty^1(\mathcal{X})$;
- (ii) $\left(c^{-1}\gamma \text{Div}(z^{(t)})\right)_{t \in \mathbb{N}}$ converges to $c^{-1}\gamma \text{Div}(\hat{z}) = (\text{Id} - \text{prox}_{c^{-1}\gamma \|\cdot\|_{\text{TV}}})(u)$.

The proof of this lemma can be found in our Report [15].

The iteration proposed in (30) to compute the proximity operator of the TV-norm is different from the projection algorithm of [6]. A similar iteration was proposed in [7] and in some other articles. The proof we gave is however simpler as it uses known properties of proximity operators. Note that computing $\text{prox}_{\|\cdot\|_{\text{TV}}}$ amounts to solving a discrete ROF-denoising. Our iteration to solve this problem is one possibility among others, see e.g. a recent report [4].

A crucial property of the D-R scheme (22) is its robustness to numerical errors that may occur when computing the proximity operators prox_{ψ} and prox_{ϕ} , see [10]. More precisely, let $a_t \in \ell^2(I)$ be an error term that models the inexact computation of $\text{prox}_{\gamma\phi}$ in (24), as the latter is obtained through (30). If the sequence of error terms $(a_t)_{t \in \mathbb{N}}$ and stepsizes $(\mu_t)_{t \in \mathbb{N}}$ in Theorem 2 obey $\sum_{t \in \mathbb{N}} \mu_t \|a_t\| < +\infty$, then the D-R algorithm (22) converges [10, Corollary 6.2]. In our experiments, using 200 inner iterations in (30) is sufficient to satisfy this requirement.

3.3 Bias Correction to Recover the Sought-After Image

Recall from (4) that $u_0 = \log S_0$ and set $\hat{u} = \widetilde{W}\hat{x}^{(N_{\text{DR}})}$ as the estimator of u_0 , where N_{DR} is the number of D-R iterations in (22). 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 $\mathbb{E}[\hat{S}] = S_0$. Expanding \hat{S} in the neighborhood of $\mathbb{E}[\hat{u}]$, we have

$$\mathbb{E}[e^{\hat{u}}] = \exp(\mathbb{E}[\hat{u}]) (1 + \text{Var}[\hat{u}]/2 + R_2) = S_0 \exp(-b_{\hat{u}}) (1 + \text{Var}[\hat{u}]/2 + R_2), \quad (33)$$

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, hence $R_2 \approx 0$. Then $b_{\hat{u}} \approx \log(1 + \text{Var}[\hat{u}]/2)$ ensures unbiasedness. Consequently, finite sample (nearly) unbiased estimates of u_0 and S_0 are respectively $\hat{u} + \log(1 + \text{Var}[\hat{u}]/2)$, and $\exp(\hat{u}) (1 + \text{Var}[\hat{u}]/2)$. $\text{Var}[\hat{u}]$ can be reasonably estimated by $\psi_1(K)$, the variance of the noise n in (4) being given in (1). Thus, given the restored log-image \hat{u} , our denoised image read:

$$\hat{S} = \exp(\hat{u}) (1 + \psi_1(K)/2). \quad (34)$$

4 Full Algorithm to Suppress Multiplicative Noise

Piecing together Lemmas 1 and 2, and Theorem 2, we write down the full multiplicative noise removal algorithm:

Task: Denoise an image S corrupted 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), step-sizes $\mu_t \in (0, 2)$, $0 < \beta_t < 1/4$ 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) $\mathcal{T}_S^{\gamma\lambda_i}(z) = \left(\max \{0, z[i] - \gamma\lambda_i \text{sign}(z[i])\} \right)_{i \in I}$, $\forall z \in \mathbb{R}^{\#I}$.

(b) $\forall (i, j) \in \Omega$, $P_{B_\infty^1(\mathcal{X})}^{-1}(z)[i, j] = \begin{cases} z[i, j] & \text{if } |z[i, j]| \leq 1 \\ z[i, j]/|z[i, j]| & \text{else.} \end{cases}$

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

(d) $\psi_1(\cdot)$ defined according to (1) (built-in Matlab function).

Initialization: Compute $v = \log S$ and transform coefficients $y = Wv$. Hard-threshold y at T to get $y_{\mathcal{T}_H}$. Choose $x^{(0)}$.

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** $N_{\text{FB}} - 1$

$$z^{(s+1)} = P_{B_\infty^1(\mathcal{X})}^{-1} \left(z^{(s)} + \beta_t \ddot{\nabla} \left(\text{Div}(z^{(s)}) - \frac{c}{\gamma} u^{(t)} \right) \right).$$

(3) Set $z^{(t)} = z^{(N_{\text{FB}})}$ and compute $w^{(t)} = c^{-1}\gamma \text{Div}(z^{(t)})$.

(4) Forward curvelet transform: $\alpha^{(t)} = Ww^{(t)}$.

(5) Compute $r^{(t)} = \text{rprox}_{\gamma\phi}(x^{(t)}) = x^{(t)} - 2\alpha^{(t)}$.

(6) Find $q^{(t)} = (\text{rprox}_{\gamma\psi} \circ \text{rprox}_{\gamma\phi}) x^{(t)} = 2 \left(y_{\mathcal{T}_H}[i] + \mathcal{T}_S^{\gamma\lambda_i}(r^{(t)}[i] - y_{\mathcal{T}_H}[i]) \right)_{i \in I} - r^{(t)}$.

(7) Update $x^{(t+1)}$: $x^{(t+1)} = (1 - \mu_t/2)x^{(t)} + (\mu_t/2)q^{(t)}$.

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

5 Experiments

In all experiments, 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 $N_{\text{DR}} = 50$. 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 [2] minimizing the criterion in (10), and the Stein-block denoising method [8] 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 various additive bounded noises. We also tried the L2-TV method where the restored log-image \hat{u} minimizes (11) and the denoised image \hat{S} involves the bias correction (34). Thanks to the bias correction, it can be seen as an improved version of the first method proposed in the recent Report [26, § 4.1]. 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 two images: Lena and Boat, all of size 256×256 and 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 (2)-(3) for $K = 10$. The running time of our denoising method is 1 minute 3 seconds for 50 iterations on Intel 2.5 GHz Core Duo. The denoising performance of any algorithm is measured in terms of peak signal-to-noise ratio (PSNR) and mean absolute-deviation MAE, namely

$$\text{PSNR} = 20 \log_{10} \sqrt{N} \|S_0\|_{\infty} / \left\| \hat{S} - S_0 \right\|_2 \text{ dB} \quad \text{and} \quad \text{MAE} = \left\| \hat{S} - S_0 \right\|_1 / N .$$

The results are depicted in Figs. 3 and 4. Note that the AA algorithm tends to over-regularize the solution. Our denoiser clearly outperforms its competitors.

References

1. Achim, A., Tsakalides, P., Bezerianos, A.: Sar image denoising via bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. Remote Sens.* 41(8), 1773–1784 (2003)
2. Aubert, G., Aujol, J.-F.: A variational approach to remove multiplicative noise. *J. on Applied Mathematics* 68(4), 925–946 (2008)
3. Aubert, G., Kornprobst, P.: *Mathematical problems in image processing*, 2nd edn. Springer, Berlin (2006)
4. Aujol, J.-F.: Some algorithms for total variation based image restoration. Report CLMA 2008-05 (2008)
5. Candès, E.J., Guo, F.: New multiscale transforms, minimum total variation synthesis. Applications to edge-preserving image reconstruction. *Signal Processing* 82 (2002)
6. Chambolle, A.: An algorithm for total variation minimization and application. *J. of Mathematical Imaging and Vision* 20(1) (2004)
7. Chambolle, A.: Total variation minimization and a class of binary MRF models. In: Rangarajan, A., Vemuri, B.C., Yuille, A.L. (eds.) *EMMCVPR 2005*. LNCS, vol. 3757, pp. 136–152. Springer, Heidelberg (2005)
8. Chesneau, C., Fadili, J., Starck, J.-L.: Stein block thresholding for image denoising. Technical report (2008)
9. Coifman, R.R., Sowa, A.: Combining the calculus of variations and wavelets for image enhancement. *Applied and Computational Harmonic Analysis* 9 (2000)
10. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization* 53(5) (2004)
11. Combettes, P.L., Pesquet, J.-C.: A Douglas-Rachford splitting approach to non-smooth convex variational signal recovery. *IEEE J. of Selected Topics in Signal Processing* 1(4), 564–574 (2007)
12. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81(3), 425–455 (1994)
13. Durand, S., Nikolova, M.: Denoising of frame coefficients using l1 data-fidelity term and edge-preserving regularization. *SIAM J. on Multiscale Modeling and Simulation* 6(2), 547–576 (2007)
14. Durand, S., Froment, J.: Reconstruction of wavelet coefficients using total variation minimization. *SIAM J. on Scientific Computing* 24(5), 1754–1767 (2003)

15. Durand, S., Fadili, J., Nikolova, M.: Multiplicative noise removal using L1 fidelity on frame coefficients. Report CMLA n.2008-40 (2008)
16. Fukuda, S., Hirose, H.: Suppression of speckle in synthetic aperture radar images using wavelet. *Int. J. Remote Sens.* 19(3), 507–519 (1998)
17. Krissian, K., Westin, C.-F., Kikinis, R., Vosburgh, K.G.: Oriented speckle reducing anisotropic diffusion. *IEEE Trans. on Image Processing* 16(5), 1412–1424 (2007)
18. Ma, J., Plonka, G.: Combined Curvelet Shrinkage and Nonlinear Anisotropic Diffusion. *IEEE Trans. on Image Processing* 16(9), 2198–2206 (2007)
19. Malgouyres, F.: Mathematical analysis of a model which combines total variation and wavelet for image restoration. *J. of information processes* 2(1), 1–10 (2002)
20. Moreau, J.-J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. *CRAS Sér. A Math*
21. Nikolova, M.: Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM J. on Numerical Analysis* 40(3), 965–994 (2002)
22. Nikolova, M.: Weakly constrained minimization. Application to the estimation of images and signals involving constant regions. *J. of Mathematical Imaging and Vision* 21(2), 155–175 (2004)
23. Pizurica, A., Wink, A.M., Vansteenkiste, E., Philips, W., Roerdink, J.B.T.M.: A review of wavelet denoising in mri and ultrasound brain imaging. *Current Medical Imaging Reviews* 2(2), 247–260 (2006)
24. Rudin, L., Lions, P.-L., Osher, S.: Multiplicative denoising and deblurring: Theory and algorithms. In: Osher, S., Paragios, N. (eds.), pp. 103–119. Springer, Heidelberg (2003)
25. Rudin, L., Osher, S., Fatemi, C.: Nonlinear total variation based noise removal algorithm. *Physica* 60D, 259–268 (1992)
26. Shi, J., Osher, S.: A nonlinear inverse scale space method for a convex multiplicative noise model. In: *UCLA 2007* (2007)
27. Ulaby, F., Dobson, M.C.: *Handbook of Radar Scattering Statistics for Terrain*. Artech House, Norwood (1989)
28. Walessa, M., Datcu, M.: Model-based despeckling and information extraction from sar images. *IEEE Trans. Geosci. Remote Sens.* 38(9), 2258–2269 (2000)
29. Welk, M., Steidl, G., Weickert, J.: Locally analytic schemes: A link between diffusion filtering and wavelets shrinkage. *Applied and Computational Harmonic Analysis* 24, 195–224 (2008)
30. Xie, H., Pierce, L.E., Ulaby, F.T.: SAR speckle reduction using wavelet denoising and markov random field modeling. *IEEE Trans. Geosci. Remote Sensing* 40(10), 2196–2212 (2002)
31. Yu, Y., Acton, S.T.: Speckle reducing anisotropic diffusion. *IEEE Trans. on Image Processing* 11(11), 1260–1270 (2002)