Removing multiplicative noise by Douglas-Rachford splitting methods

G. Steidl and T. Teuber [*]

May 19, 2009

Abstract

Multiplicative noise appears in various image processing applications, e.g., in synthetic aperture radar (SAR), ultrasound imaging or in connection with blur in electronic microscopy, single particle emission computed tomography (SPECT) and positron emission tomography (PET). In this paper, we consider a variational restoration model consisting of the I-divergence as data fitting term and the total variation semi-norm or nonlocal means as regularizer. Although the I-divergence is the typical data fitting term when dealing with Poisson noise we substantiate why it is also appropriate for cleaning Gamma noise. We propose to compute the minimizer of our restoration functional by applying Douglas-Rachford splitting techniques, resp. alternating split Bregman methods, combined with an efficient algorithm to solve the involved nonlinear systems of equations. We prove the Q-linear convergence of the latter algorithm. Finally, we demonstrate the performance of our whole scheme by numerical examples. It appears that the nonlocal means approach leads to very good qualitative results.

1 Introduction

We are interested in restoring images $f : \Omega \to \mathbb{R}$ arising from original images u by corruption with (uncorrelated) multiplicative noise η of mean 1, i.e.,

$$f = u\eta. \tag{1}$$

The task of removing multiplicative noise appears in many applications, e.g., in synthetic aperture radar (SAR), where the noise is assumed to follow a Gamma distribution, in ultrasound imaging, where we are confronted with Rayleigh distributed [41], K-distributed [14, 24, 25] or Rician distributed [23] noise. Moreover, in electronic microscopy [27, 32], single particle emission computed tomography (SPECT) [32] and positron emission tomography (PET) [36], multiplicative Poisson noise appears in connection with blur.
In this paper, we focus on Gamma distributed noise although our model is appropriate for Poisson distributed noise as well. Recently, various variational models for removing

---

[*]University of Mannheim, Dept. of Mathematics and Computer Science

Gamma noise were proposed. Following the MAP estimator for multiplicative Gamma noise, Aubert and Aujol [3] introduced a non-convex model whose data term was subsequently adopted in a convex model by Shi and Osher in [39]. Indeed these authors considered a more general data fitting term, which includes also the model in [34]. They applied a corresponding relaxed inverse scale space flow as denoising technique. The model of Osher and Shi was modified in [22] by adding a quadratic term to get a simpler alternating minimization algorithm. A variational model involving curvelet coefficients for cleaning multiplicative Gamma noise was considered in [13].

Beyond variational approaches there exist other techniques to remove multiplicative noise, e.g., local linear minimum mean square error approaches [29, 28] or anisotropic diffusion methods [45, 1, 26] which will not be addressed in this paper.

In the above variational models the data fitting term arises from the MAP estimator for multiplicative Gamma noise. However, in deblurring problems, where we have frequently Poisson noise, Csiszár's I-divergence [11] is usually applied as data fitting term. For the expectation-maximization (EM) approach related to the I-divergence model in deblurring problems see [31, 33] and the references therein and for the EM - total variation (TV) model we refer to [32, 36].

In this paper, we consider an I-divergence - TV and I-divergence - nonlocal means (NL-means) model for denoising. We motivate why the I-divergence data fitting term typically used in the context of Poisson noise is also appropriate when dealing with Gamma noise. We develop iterative algorithms for computing the minimizers of our functionals by applying Douglas-Rachford splitting techniques. Such methods were first applied in image processing in [9]. Note that for our setting the Douglas-Rachford splitting is equivalent to an alternating split Bregman algorithm [37, 38].

This paper is organized as follows: We start by reviewing some variational denoising methods in Section 2. In particular, we consider their performance for two-pixel signals. We will how the minimizer of the functional proposed in [39] is related to the minimizer of our I-divergence - TV model. In Section 3 we propose to compute the minimizers of our functionals by applying Douglas-Rachford splitting combined with an efficient algorithm to solve the involved nonlinear systems of equations. Moreover, we prove the Q-linear convergence of the latter semi-implicit schemes. Finally, we demonstrate the performance of our algorithms for the I-divergence - TV and the I-divergence - NL-means model by numerical examples in Section 4. Conclusions are given in Section 5.

2   Edge preserving variational methods for removing multiplicative noise

Variational methods aim to restore the original image by finding the minimizer of some appropriate functional

$$E(u) := \Psi(u) + \lambda\Phi(u), \qquad \lambda > 0,$$

where $\Psi = \Psi_f$ denotes the *data fitting term* and Φ is a *regularization term* which includes prior information about the original image.

In general, the data fitting term is deduced by maximizing the a-posteriori probability density $p(u|f)$ (MAP estimation). Most papers deal with additive noise, i.e., $f = u + \eta$. If u is corrupted by additive white Gaussian noise, this leads to the data fitting term $\Psi_f(u) := \int_\Omega (f - u)^2 \, dx$. A frequently applied regularization term is the total variation (TV) semi-norm suggested by Rudin, Osher and Fatemi (ROF) [35] $|u|_{BV} := \sup_{p \in C_0^1, \| \, |p| \, \|_\infty \leq 1} \int_\Omega u \operatorname{div} p \, dx$ which is formally (for sufficiently regular u)

$$\Phi(u) = \int_\Omega |\nabla u| \, dx. \tag{2}$$

In the case of additive Gaussian noise, the minimizer $\hat{u}$ of the whole ROF functional

$$\frac{1}{2} \int_\Omega (f - u)^2 \, dx + \lambda \int_\Omega |\nabla u| \, dx \tag{3}$$

has many desirable properties. It preserves important structures such as edges, fulfills a *maximum-minimum principle* which reads in the discrete n- pixel setting as $f_{min} \leq \hat{u}_i \leq f_{max}$, where f_{min} and f_{max} denote the minimal and maximal coefficient of f, resp., and preserves the mean value, i.e.,

$$\sum_{i=1}^n \hat{u}_i = \sum_{i=1}^n f_i.$$

The drawback of model (3) consists in its staircasing effect so that meanwhile various alternative regularizers were considered. Among them, the NL-means regularization term leads to very good denoising results. The idea of nonlocal means goes back to [6] and was incorporated into the variational framework in [17, 19, 18, 46]. We refer to these papers for further information on NL-means. Based on some pre-computed weights w the regularization term is given by

$$\Phi(u) = \int_\Omega |\nabla_w u| \, dx, \quad |\nabla_w u| := \left(\int_\Omega (u(y) - u(x))^2 w(x, y) \, dy \right)^{1/2}. \tag{4}$$

In the following, we review variational methods for removing multiplicative noise, where we restrict our attention to TV regularizers. To see the differences between the models it is sometimes useful to apply them to the simplest signals $f := (f_1, f_2)^{\mathrm{T}}$ consisting of only two pixels so that (3) becomes

$$\Psi(f_1, u_1) + \Psi(f_2, u_2) + \lambda |u_2 - u_1|. \tag{5}$$

log-model. By (1) it seems to be more appropriate to include quotients rather than differences of f and u into the fitting term, e.g., $\max\{\frac{f}{u}, \frac{u}{f}\}$. Taking the logarithm of this term and setting $w := \log u$ we get $|w - \log f|$ and using w in the regularization term (2) we obtain, for a noisy signal $f > 0$, the log-model

$$\hat{w} := \operatorname*{argmin}_{w \in BV} \left\{ \frac{1}{2} \int (w - \log f)^2 \, dx + \lambda \int |\nabla w| \, dx \right\}, \quad \hat{u} = e^{\hat{w}}. \tag{6}$$

3

This is the usual ROF-model (3) for w and $\log f$. Therefore the maximum-minimum principle carries directly over to $\hat{u}$. However, the mean value preservation $\sum_{i=1}^{n} \hat{w}_i = \sum_{i=1}^{n} \log f_i$ leads to

$$\prod_{i=1}^{n} \hat{u}_i = \prod_{i=1}^{n} f_i.$$

This means that the *log-model preserves the geometric mean rather than the arithmetic mean*. For example, if λ is large enough, then $\hat{u}_i = (\prod_{j=1}^{n} f_j)^{1/n}$ for all $i = 1, \ldots, n$ which is indeed smaller than the mean of f provided that f is not the constant signal. So this property is a severe problem if one wants to use such an approach with a strong multiplicative noise since in this case the mean of the restored image is much less than the one of the original image. Such a model can therefore not be considered as a good one for multiplicative noise removal. We want to have a look at the two pixel model (5) for the setting (6).

Example (two-pixel signals). We assume that $f_1 \geq f_2 > 0$. Setting the gradient with respect to w_i, $i = 1, 2$ to zero we obtain that the minimizer $\hat{u}_1, \hat{u}_2$ move to $\sqrt{f_1 f_2}$ with increasing λ as follows:

$$\hat{u}_1 = f_1 e^{-\lambda}, \ \hat{u}_2 = f_2 e^{\lambda} \qquad \text{for} \quad 0 \leq \lambda < \frac{1}{2} \log \frac{f_1}{f_2},$$

$$\hat{u}_1 = \hat{u}_2 = \sqrt{f_1 f_2} \qquad \text{for} \quad \frac{1}{2} \log \frac{f_1}{f_2} \leq \lambda.$$

AA-model. Based on the MAP estimator for multiplicative Gamma noise, Aubert and Aujol [3] proposed to determine the denoised image as a minimizer in $\{u \in BV : u > 0\}$ of the following, in general non-convex, functional

$$\int_{\Omega} \frac{f}{u} + \log u \, dx + \lambda \int_{\Omega} |\nabla u| \, dx. \tag{7}$$

While the data fitting term follows canonically from the MAP approach related to the Gamma distribution, the choice of the regularization term is flexible and we will see in the following that $|\nabla \log u|$ seems to be a better choice. In particular, it was observed in numerical examples [3, 39] that the noise survives much longer at low image values if we increase the regularization parameter. This is also indicated by our simple two pixel model.

Example (two-pixel signals). We restrict our attention to the case $0 < f_2 \leq f_1 \leq 3 f_2$. This may appear if f_1 and f_2 are disturbed versions of a constant function $u_1 = u_2 = u$, i.e., $f_1 = (1 + \nu) u$ and $f_2 = (1 - \nu) u$, where $0 \leq \nu \leq 1/2$. Then the minimizer reads

$$\hat{u}_1 = \frac{-1 + \sqrt{1 + 4\lambda f_1}}{2\lambda}, \ \hat{u}_2 = \frac{1 - \sqrt{1 - 4\lambda f_1}}{2\lambda} \qquad \text{for} \quad 0 < \lambda < \frac{2(f_1 - f_2)}{(f_1 + f_2)^2},$$

$$\hat{u}_1 = \hat{u}_2 = \frac{f_1 + f_2}{2} \qquad \text{for} \quad \frac{2(f_1 - f_2)}{(f_1 + f_2)^2} \leq \lambda.$$

Assuming as above that $f_i = (1 - (-1)^i \nu)u$, we see that we have to choose $\lambda \geq \frac{\nu}{u}$ to get the original constant signal u. This means that λ must be chosen larger for smaller values of u.

SO-model. Shi and Osher [39] suggested to keep the data fitting term in (7) but to replace the regularizer $|\nabla u|$ by $|\nabla \log u|$. Moreover, setting as in the log-model $w := \log u$, this results in the convex functional

$$\hat{w} = \underset{w \in BV}{\mathrm{argmin}} \left\{ \int_\Omega f e^{-w} + w \, dx + \lambda \int_\Omega |\nabla w| \, dx \right\}, \quad \hat{u} = e^{\hat{w}} \tag{8}$$

which overcomes the drawback of (7) as we will see by looking at our two-pixel model (5).

Example (two-pixel signals). Let $f_1 \geq f_2 > 0$. Then the solution of (5) is

$$\hat{u}_1 = \frac{f_1}{1+\lambda}, \; \hat{u}_2 = \frac{f_2}{1-\lambda} \qquad \text{for} \quad 0 \leq \lambda < \frac{f_1 - f_2}{f_1 + f_2},$$

$$\hat{u}_1 = \hat{u}_2 = \frac{f_1 + f_2}{2} \qquad \text{for} \quad \frac{f_1 - f_2}{f_1 + f_2} \leq \lambda.$$

For $f_i = (1 - (-1)^i \nu)u$, the value u is reconstructed if $\lambda \geq \nu$ which is independent of the size of u.

Indeed, Shi and Osher considered a more general approach with data fitting term $a f e^{-w} + \frac{b}{2} f^2 e^{-2w} + (a + b)w$ which includes also the model in [34], but $b \neq 0$ gives in general no better results. Besides, the authors computed the corresponding relaxed inverse scale space flow to further improve the quality of the restored image.

I-divergence model. In connection with deblurring in the presence of multiplicative noise the *I-divergence*, also called *generalized Kullback-Leibler* divergence

$$I(f, u) := \int_\Omega f \log \frac{f}{u} - f + u \, dx$$

is typically used as data fitting term. The I-divergence is the *Bregman distance* [5] of the function $F(u) := \int_\Omega u \log u - u \, dx$, i.e., $I(f, u) = F(f) - F(u) - \langle p, f - u \rangle$, where $p \in \partial F(u)$. Therefore it shares the useful properties of a Bregman distance, in particular $I(f, u) \geq 0$. Ignoring the constant terms, the corresponding convex denoising model reads

$$\hat{u} = \underset{u \in BV, u > 0}{\mathrm{argmin}} \left\{ \int_\Omega u - f \log u \, dx + \lambda \int_\Omega |\nabla u| \, dx \right\}. \tag{9}$$

Having the MAP approach in mind, this model seems to be better related to Poisson noise than to Gamma noise. This may be the reason why it was not considered for denoising, e.g., of SAR images up to now. But having a closer look at the model, we see that the gradient of the data fitting terms in (9) and (8) coincide if we use again the relation $\log \hat{u} = \hat{w}$. Moreover, if we add TV-regularization, then both functionals have the same minimizer. This can be seen roughly as follows:

Since $\nabla e^w = e^w \nabla w$ we have for $u = e^w$ that $\nabla u(x) = 0$ if and only if $\nabla w(x) = 0$. The minimizers $\hat{w}$ and $\hat{u}$ of the functionals (8) and (9) are unique and given by

$$
\begin{aligned}
0 &= 1 - f e^{-\hat{w}} - \lambda \operatorname{div} \frac{\nabla \hat{w}}{|\nabla \hat{w}|} \quad \text{for} \quad |\nabla \hat{w}(x)| \neq 0, \\
0 &= 1 - \frac{f}{\hat{u}} - \lambda \operatorname{div} \frac{\nabla \hat{u}}{|\nabla \hat{u}|} \quad \text{for} \quad |\nabla \hat{u}(x)| \neq 0.
\end{aligned}
$$

Since $\frac{\nabla w}{|\nabla w|} = \frac{e^w \nabla w}{e^w |\nabla w|} = \frac{\nabla u}{|\nabla u|}$, we obtain the assertion.

A sophisticated proof may be given via subgradient inclusions. We will use this approach for the discrete setting in Proposition 3.7, where we have due to the discretization of the derivative that coincidence holds only true in 1D.

3   Minimization by Douglas-Rachford splitting

In the following, we work within a discrete setting, i.e., we consider columnwise reshaped images $f : \mathbb{R}^n \to \mathbb{R}$. Products, quotients etc. of vectors are meant componentwise. By $D : \mathbb{R}^n \to \mathbb{R}^{mn}$ we denote either

i) some discretization of the gradient operator as, e.g., those in [7, 40] with $m = 2$, see (26), or

ii) the NL-means operator with binary weights introduced in [19] with m associated to the number of permitted neighbors, see Section 4. Note that as in i) the rows of D contain exactly one entry -1 and one entry 1 or are zero rows.

Further, for $p := (\mathrm{p}_1, \ldots, \mathrm{p}_m)^{\mathrm{T}} \in \mathbb{R}^{mn}$ with $\mathrm{p}_k := (p_{j+(k-1)n})_{j=1}^n$ we use the notation $|p| := (\mathrm{p}_1^2 + \ldots + \mathrm{p}_m^2)^{1/2} \in \mathbb{R}^n$.
We ask for the minimizer $\hat{u}$ of the discrete model

$$
\min_{u \in \mathbb{R}^n} \left\{ \Psi(u) + \lambda \Phi(Du) \right\}, \tag{10}
$$

with

$$
\Psi(u) := \begin{cases} \langle 1, u - f \log u \rangle & \text{if } u > 0, \\ \infty & \text{otherwise} \end{cases} \quad \text{and} \quad \Phi(p) := \| \, |p| \, \|_1. \tag{11}
$$

If D is given by i) then we refer to (10) as I-divergence - TV model and if D is determined by ii) we call it I-divergence - NL-means model. The functional in (10) is proper, lower semi-continuous, coercive and strictly convex on its domain. Therefore there exists a unique minimizer. The dual problem of (10) reads

$$
- \min_{p \in \mathbb{R}^{mn}} \left\{ \Psi^*(-D^*p) + \lambda \Phi^*(\lambda^{-1} p) \right\} \tag{12}
$$

with the conjugate functions

$$
\Psi^*(v) = \begin{cases} -\langle f, \log(1-v) \rangle + \langle f, \log f - 1 \rangle & \text{if } v < 1, \\ \infty & \text{otherwise} \end{cases} \quad \text{and} \quad \Phi^*(p) = \iota_C(p),
$$

6

where

$$\iota_C(p) := \begin{cases} 0 & \text{if } p \in C, \\ \infty & \text{otherwise} \end{cases}$$

denotes the *indicator function* of $C := \{p \in \mathbb{R}^{mn} : \| |p| \|_\infty \leq 1\}$. There is no duality gap, i.e., (10) and (12) take the same value and the minimizers are related by

$$\hat{u} = f/(1 + D^*\hat{p}). \tag{13}$$

The following proposition describes properties of the minimizer of (10).

Proposition 3.1. *The solution $\hat{u}$ of (10) has the following properties:*

i) Minimum-maximum principle:

$$f_{min} \leq \hat{u}_i \leq f_{max} \quad \text{for all} \quad i = 1, \ldots, n,$$

where f_{min} and f_{max} denote the values of the smallest and largest coefficient of f.

ii) Averaging property:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{f_i}{\hat{u}_i} = 1.$$

The second property is desirable by (1) and since the mean of η is 1.

Proof: i) The first property follows in the same way as in [3, Theorem 4.1]. We have only to verify the relations

$$J\big(\min(u, f_{max})\big) \leq J(u), \quad J\big(\max(u, f_{min})\big) \leq J(u),$$

where $J(u) := \Phi(Du)$. By the structure of Φ and D, we see that $J(u)$ contains only summands of the form $(u_i - u_j)^2$. Thus it remains to show that

$$|u_i - u_j| \geq |\min(u_i, f_{max}) - \min(u_j, f_{max})|,$$
$$|u_i - u_j| \geq |\max(u_i, f_{min}) - \max(u_j, f_{min})|.$$

The case $u_i = u_j$ is trivial so that we assume $u_i > u_j$. If $f_{max} \geq u_i$ or $u_j \geq f_{max}$, resp. $f_{min} \geq u_i$ or $u_j \geq f_{min}$ we are done. For $u_i > f_{max} > u_j$, the first inequality becomes $|u_i - u_j| \geq |f_{max} - u_j|$ which is true since $f_{max} > 0$. Similarly, we get for $u_i > f_{min} > u_j$ the correct inequality $|u_i - u_j| \geq |u_i - f_{min}|$.
ii) The second property follows from (13) and since $1^\mathrm{T} D^* = 0$. Namely, we have

$$\frac{1}{n} \sum_{i=1}^{n} \frac{f_i}{\hat{u}_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{f_i}{f_i/(1 + (D^*\hat{p})_i)} = 1 + \frac{1}{n} \sum_{i=1}^{n} (D^*\hat{p})_i = 1. \quad \square$$

By Fermat's rule we know that $\hat{p}$ is a minimizer of (12) if and only if

$$0 \in \partial \left(\Psi^* \circ (-D^*) \right)(\hat{p}) + \partial \Phi^* (\lambda^{-1} \hat{p}). \tag{14}$$

Since Ψ^* and Φ^* are proper, closed and convex, the operators $\partial \left(\Psi^* \circ (-D^*) \right)$ and $\partial \Phi^*$ are maximal monotone, see [4]. The second operator is indeed set-valued. For a maximal monotone operator A, the *resolvent* $J_A := (I + A)^{-1}$ of A is single-valued and firmly non-expansive, see [4]. Inclusions of the form (14) can be solved by various splitting techniques like *forward-backward splitting* or *Douglas-Rachford splitting* (DRS), see [30, 15]. In this paper, we focus on the DRS because it leads to an efficient algorithm. Note that DRS was first considered in [12] for linear operators.

Theorem 3.2. *Let H be a Hilbert space and $A, B : H \to 2^H$ maximal monotone operators. Assume that a solution $\hat{p}$ of*

$$0 \in A(p) + B(p)$$

exists. Then, for any initial elements $t^{(0)}$ and $p^{(0)}$ and any $\eta > 0$, the following DRS algorithm converges weakly to an element $\hat{t}$:

$$
\begin{aligned}
t^{(k+1)} &= J_{\eta A}(2p^{(k)} - t^{(k)}) + t^{(k)} - p^{(k)}, \\
p^{(k+1)} &= J_{\eta B}(t^{(k+1)}).
\end{aligned}
$$

Furthermore, it holds that $\hat{p} := J_{\eta B}(\hat{t})$ satisfies $0 \in A(\hat{p}) + B(\hat{p})$. If H is finite dimensional, then the sequence $\left(p^{(k)} \right)_{k \in \mathbb{N}}$ converges to $\hat{p}$.

For the proof see, e.g., [10]. Note that the iterates in the DRS must not be computed exactly. Their convergence is also ensured if we allow summable errors [15].

By the following Proposition 3.3, the DRS applied to the dual problem (14) coincides with the so-called alternating split Bregman algorithm for the primal problem. To introduce this algorithm we consider the equivalent problem of (10)

$$\min_{u \in \mathbb{R}^n, d \in \mathbb{R}^{mn}} \underbrace{\left\{ \Psi(u) + \lambda \Phi(d) \right\}}_{E(u,d)} \quad \text{subject to} \quad d = Du$$

and apply the *augmented Lagrangian method* to compute the minimizer iteratively by

$$
\begin{aligned}
(u^{(k+1)}, d^{(k+1)}) &= \operatorname*{argmin}_{u \in \mathbb{R}^n, d \in \mathbb{R}^{mn}} \left\{ E(u,d) + \langle b^{(k)}, Du - d \rangle + \frac{1}{2\gamma} \| Du - d \|_2^2 \right\} \\
&= \operatorname*{argmin}_{u \in \mathbb{R}^n, d \in \mathbb{R}^{mn}} \left\{ E(u,d) + \frac{1}{2\gamma} \| \gamma\, b^{(k)} + Du - d \|_2^2 \right\}, \\
b^{(k+1)} &= b^{(k)} + \frac{1}{\gamma}(Du^{(k+1)} - d^{(k+1)}), \qquad \gamma > 0.
\end{aligned}
$$

If we replace $b^{(k)}$ by $b^{(k)}/\gamma$, this method is also known in image processing as *split Bregman algorithm*, see [16, 20, 44, 42]. Since the first functional is in general hard to minimize,

one uses instead the following *alternating split Bregman algorithm*:

$$u^{(k+1)} = \operatorname*{argmin}_{u \in \mathbb{R}^n} \left\{ \Psi(u) + \frac{1}{2\gamma} \|b^{(k)} + Du - d^{(k)}\|_2^2 \right\}, \tag{15}$$

$$d^{(k+1)} = \operatorname*{argmin}_{d \in \mathbb{R}^{mn}} \left\{ \Phi(d) + \frac{1}{2\gamma\lambda} \|b^{(k)} + Du^{(k+1)} - d\|_2^2 \right\}, \tag{16}$$

$$b^{(k+1)} = b^{(k)} + Du^{(k+1)} - d^{(k+1)}. \tag{17}$$

The convergence of this algorithm is ensured by Theorem 3.2 and the following proposition from [38].

Proposition 3.3. *The alternating split Bregman algorithm coincides with the DRS algorithm applied to the dual problem with $A := \partial(\Psi^* \circ (-D^*))$ and $B := \partial \Phi^*(\lambda^{-1} \cdot)$, where $\eta = 1/\gamma$ and*

$$t^{(k)} = \eta(b^{(k)} + d^{(k)}), \quad p^{(k)} = \eta b^{(k)}, \quad k \geq 0. \tag{18}$$

Moreover, the existence of a unique minimizer of (10) guarantees the convergence of $\{u^{(k)}\}_{k \in \mathbb{N}}$ defined in (15) to this minimizer in our finite dimensional setting, see [37].

Let us consider the minimization problems (15) and (16) appearing in the alternating split Bregman algorithm in more detail. By the simple structure of Φ in (11), the solution of (16) is given by

$$d^{(k+1)} = T_{\gamma\lambda}(b^{(k)} + Du^{(k+1)}),$$

where T_τ denotes the *coupled shrinkage* function which is determined componentwise for $p = (p_1, \ldots, p_m)^T \in \mathbb{R}^{mn}$ with $p_k := (p_{j+(k-1)n})_{j=1}^n$ by

$$T_\tau(p_{j+(k-1)n}) := \begin{cases} 0 & \text{if } |p|_j \leq \tau, \\ p_{j+(k-1)n} - \tau p_{j+(k-1)n}/|p|_j & \text{if } |p|_j > \tau, \end{cases}$$

see, e.g., [8, 43].

In contrast to the original problem (10), the functional (15) in the alternating split Bregman algorithm has a quadratic penalizer which is differentiable. Setting the gradient to zero, we see that w is a solution of (15) if $w > 0$ satisfies

$$0 = 1 - \frac{f}{w} + \frac{1}{\gamma}\left(D^*Dw + D^*(b^{(k)} - d^{(k)})\right), \tag{19}$$

$$0 = \gamma w - \gamma f + w\, D^*Dw - w\, D^*(d^{(k)} - b^{(k)}). \tag{20}$$

This nonlinear system of equations can be solved in various ways, e.g., by Newton- or Newton-like methods if a good initial guess exists or by applying DRS again as in the following remark, see also [9].

Remark 3.4. *We apply DRS in (19) with*

$$A(w) := \frac{1}{\gamma}\left(D^*Dw + D^*(b^{(k)} - d^{(k)})\right), \quad B(w) := 1 - \frac{f}{w}$$

and obtain

$$v \;=\; (I + \eta A)w = w + \frac{\eta}{\gamma}\big(D^*Dw + D^*(b^{(k)} - d^{(k)})\big),$$

$$\Leftrightarrow \quad w \;=\; J_{\eta A}(v) = (I + \frac{\eta}{\gamma}\,D^*D)^{-1}\big(v - \frac{\eta}{\gamma}\,D^*(b^{(k)} - d^{(k)})\big)$$

and

$$v = (I + \eta B)w = w + \eta\big(1 - \frac{f}{w}\big) \quad \Leftrightarrow \quad w = J_{\eta B}(v) = \big(v - \eta + \sqrt{(v - \eta)^2 + 4\eta f}\,\big)/2.$$

Note that this formula guarantees that w is positive. However, the convergence of the algorithm is rather slow.

We propose to solve (20) by an efficient method which can be deduced directly from (20) by adding τw for some $\tau > 0$ to both sides of the equation and using a semi-implicit iterative version. Let $\Lambda_{w^{(k)}} := \mathrm{diag}\big((w_j^{(k)})_{j=1}^n\big)$.

Initialization: $w^{(0)} := u^{(k)}$
For $j = 0, 1, \ldots$ solve until a stopping criterion is reached

$$\big((\tau + \gamma)I + \Lambda_{w^{(j)}}\, D^*D\big)\, w^{(j+1)} := \gamma f + w^{(j)}\big(D^*(d^{(k)} - b^{(k)}) + \tau\, 1\big).$$

By the following theorem, the sequence produced by this process converges for sufficiently large $\tau \geq 0$ to the solution $w > 0$ of (20). To this end, note that for our matrices D, the matrix $A := D^*D$ is a positive semidefinite *L-matrix*, i.e., all diagonal elements of A are positive while all non-diagonal elements are not. For the NL-means matrix $D = (d_1 \ldots d_n)$ this can be seen since the entries of A are $a_{ii} = \|d_i\|_2^2$ and $a_{ij} = \langle d_i, d_j \rangle$ and the positions of the entries -1 and 1 in d_i cannot match those of the same entries in d_j for $i \neq j$.

Theorem 3.5. *Let $A \in \mathbb{R}^{n,n}$ be a positive semidefinite L-matrix and let $c \in \mathbb{R}^n$ and $\gamma > 0$ be given. Then, for sufficiently large $\tau \geq 0$ and $w^{(0)} > 0$, the sequence $\{w^{(j)}\}_{j \in \mathbb{N}_0}$ produced by*

$$\big((\tau + \gamma)I + \Lambda_{w^{(j)}}\, A\big)\, w^{(j+1)} = \gamma f + w^{(j)}\,(c + \tau\, 1) \tag{21}$$

fulfills $w^{(j)} > 0$ and converges Q-linearly to the solution $w > 0$ of

$$0 = \gamma w - \gamma f + w\, Aw - w\, c. \tag{22}$$

Note that in our application c is fixed but may depend on γ.

Proof. 1. First, we show that we can obtain a componentwise positive sequence $\{w^{(j)}\}_{j \in \mathbb{N}_0}$ for sufficiently large τ. To this end, choose $\tau \geq 0$ such that the vector on right-hand side of (21) has only positive entries; take for example $\tau := -\min_j c_j$ if c has negative components and $\tau = 0$ otherwise. Since $w^{(j)} > 0$ the matrix on the left-hand side of (21) is a strictly (row) diagonal dominant *L-matrix* and therefore an *M-matrix*, i.e., the inverse matrix

exists and has only nonnegative entries. Thus, if $w^{(j)} > 0$, then the same holds for $w^{(j+1)}$, i.e., for the whole sequence $\{w^{(j)}\}_{j \in \mathbb{N}_0}$ if we start with $w^{(0)} > 0$.

2. Next, we show that $\|w - w^{(j)}\|_\infty$ decreases with j. By (22) and (21) we obtain with $A_w := \Lambda_w A$ and $A_j := \Lambda_{w^{(j)}} A$ that

$$
\begin{aligned}
\big((\tau+\gamma)I + A_w\big)\, w - \big((\tau+\gamma)I + A_j\big)\, w^{(j+1)} &= (c + \tau\, 1)\, (w - w^{(j)}), \\
(\tau+\gamma)(w - w^{(j+1)}) + A_w\, w - A_j\, w^{(j+1)} &= (c + \tau\, 1)\, (w - w^{(j)})
\end{aligned}
$$

and since

$$
\begin{aligned}
A_w\, w - A_j\, w^{(j+1)} &= A_w\, w - A_j\, w + A_j\, w - A_j\, w^{(j+1)} \\
&= (w - w^{(j)})\, Aw + A_j(w - w^{(j+1)})
\end{aligned}
$$

this can be rewritten as

$$
\big((\tau+\gamma)I + A_j\big)(w - w^{(j+1)}) = (Aw - c - \tau\, 1)\, (w^{(j)} - w).
$$

Further, we see by (22) that $Aw - c = \gamma(f/w - 1)$ so that

$$
\begin{aligned}
\big((\tau+\gamma)I + A_j\big)(w - w^{(j+1)}) &= (\gamma(f/w - 1) - \tau\, 1)\, (w^{(j)} - w), \\
\big(I + \frac{1}{\tau+\gamma} A_j\big)(w - w^{(j+1)}) &= \frac{\tau\, 1 - \gamma(f/w - 1)}{\tau+\gamma}\, (w - w^{(j)}). \qquad (23)
\end{aligned}
$$

For $\tau > \gamma\left(\frac{1}{2}\|f/w\|_\infty - 1\right)$ we obtain that

$$
\frac{\|\tau\, 1 - \gamma(f/w - 1)\|_\infty}{\tau+\gamma} < 1. \qquad (24)
$$

Since A is a positive semidefinite L-matrix, it is (row) diagonal dominant and consequently the vector y defined by

$$
(I + \frac{1}{\tau+\gamma} A_j)\, 1 = 1 + \underbrace{\frac{1}{\tau+\gamma} \Lambda_{w^{(j)}}\, A\, 1}_{y},
$$

fulfills $y \geq 0$. Hence,

$$
1 = (I + \frac{1}{\tau+\gamma} A_j)^{-1}\, 1 + (I + \frac{1}{\tau+\gamma} A_j)^{-1}\, y
$$

and regarding that $(I + \frac{1}{\tau+\gamma} A_j)^{-1}$ has only nonnegative entries, we see that the sum of the row entries of $(I + \frac{1}{\tau+\gamma} A_j)^{-1}$ is never larger than 1. Together with (23) and (24) this implies

$$
\|w - w^{(j+1)}\|_\infty \leq \alpha\, \|w - w^{(j)}\|_\infty, \qquad \alpha < 1,
$$

so that the algorithm converges Q-linearly. $\qquad \square$

11

Remark 3.6. *If the right-hand sides of the above linear systems of equations $\gamma f + u^{(k)}\big(D^*(d^{(k)} - b^{(k)})\big)$, $k = 0, 1, \dots$ are positive and $f/w - 1 \in (-1, 1)$, i.e, $0 < f/w < 2$, then we can use $\tau = 0$ in our algorithm. The latter condition is realistic since f is approximately a noisy variant of w, i.e., $(1 - \epsilon)w \le f \le (1 + \epsilon)w$, $\epsilon \in (0, 1)$.*

Finally, let us deal with the coincidence of the SO-model and our I-divergence model in the discrete setting.

Proposition 3.7. *Let $D \in \mathbb{R}^{mn,n}$ be a matrix which rows are zero rows or contain exactly one entry 1 and one entry -1. Then, in the case $m = 1$, the minimizer $\hat u$ of (10) coincides with $e^{\hat w}$, where $\hat w$ is the minimizer of the discrete SO-model*

$$\hat w := \operatorname*{argmin}_{w \in \mathbb{R}^n}\{\langle 1, fe^{-w} + w\rangle + \lambda \Phi(Dw)\}. \tag{25}$$

Proof. We have that $\hat u$ and $\hat w$ are the minimizers of (10) and (25), resp., if and only if

$$
\begin{aligned}
0 &\in 1 - \frac{f}{\hat u} + \lambda D^* \partial \Phi(D\hat u), \\
0 &\in 1 - f\, e^{-\hat w} + \lambda D^* \partial \Phi(D\hat w).
\end{aligned}
$$

If $\hat u = e^{\hat w}$, then $1 - \frac{f}{\hat u} = 1 - f\, e^{-\hat w}$. Next we have a look at the subdifferentials. It is well-known, see, e.g., [2] for the continuous case, that $v \in \partial \Phi(Dw)$ if and only if $v = D^*p$ for some $p \in \mathbb{R}^{mn}$ and

$$\langle p, Dw\rangle = \Phi(Dw), \quad \| |p| \|_\infty \le 1.$$

The equality can be rewritten as

$$\sum_{i=1}^n \sum_{j=0}^{m-1} p_{i+jn}(Dw)_{i+jn} = \sum_{i=1}^n \Big(\sum_{j=0}^{m-1}(Dw)^2_{i+jn}\Big)^{1/2}.$$

Set $d_i := ((Dw)_{i+jn})_{j=0}^{m-1}$, $i = 1, \dots, n$. Applying the Cauchy-Schwarz inequality to the inner sums on the left hand side, we see with $\| |p| \|_\infty \le 1$ that

$$\sum_{i=1}^n \big|\sum_{j=0}^{m-1} p_{i+jn}(Dw)_{i+jn}\big| \le \sum_{i=1}^n \Big(\sum_{j=0}^{m-1} p^2_{i+jn}\Big)^{1/2} \|d_i\|_2 \le \sum_{i=1}^n \|d_i\|_2$$

where equality holds true if and only if for each $i \in \{1, \dots, n\}$ one of the following settings appears

  i) $\|d_i\|_2 = 0$ and $(p_{i+jn})_{j=0}^{m-1}$ arbitrary with $\sum_{j=0}^{m-1} p^2_{i+jn} \le 1$ or

  ii) $\|d_i\|_2 \ne 0$ and $(p_{i+jn})_{j=0}^{m-1} = \alpha d_i$, $\sum_{j=0}^{m-1} p^2_{i+jn} = 1$. The last two equalities imply that $\alpha = 1/\|d_i\|_2$ so that $p_{i+jn} = (Dw)_{i+jn}/\|d_i\|_2$, $j = 0, \dots, m-1$.

12

Since the exponential function is strictly monotone, the case i) appears for $D\hat{w}$ if and only if it appears also for $De^{\hat{w}}$. In the second case, we get if the $(i+jn)$–th row of D contains 1 at column $(i+jn)_1$ and -1 at column $(i+jn)_2$ that

$$p_{i+jn} = \frac{\hat{w}_{(i+jn)_1} - \hat{w}_{(i+jn)_2}}{\left(\sum_{j=0}^{m-1}(\hat{w}_{(i+jn)_1} - \hat{w}_{(i+jn)_2})^2\right)^{1/2}}, \quad \text{for } w := \hat{w}$$

$$p_{i+jn} = \frac{e^{\hat{w}_{(i+jn)_1}} - e^{\hat{w}_{(i+jn)_2}}}{\left(\sum_{j=0}^{m-1}(e^{\hat{w}_{(i+jn)_1}} - e^{\hat{w}_{(i+jn)_2}})^2\right)^{1/2}}, \quad \text{for } w := e^{\hat{w}}.$$

If $m = 1$, then the right-hand sides are just $\text{sgn}\,(\hat{w}_{(i+jn)_1} - \hat{w}_{(i+jn)_2})$, resp., $\text{sgn}\,\left(e^{\hat{w}_{(i+jn)_1}} - e^{\hat{w}_{(i+jn)_2}}\right)$ and coincide since the exponential function is strictly monotone increasing. This finishes the proof. □

Note that the proof shows that the discrete models are in general not identical for $m \geq 2$.

4 Numerical Results

For our numerical examples we use MATLAB implementations. We restrict our attention to multiplicative noise which follows a Gamma distribution with density function

$$g(x) := \frac{L^L}{\Gamma(L)}x^{L-1}\exp(-Lx)\,1_{x\geq 0}.$$

Hence η has mean 1 and standard deviation $1/\sqrt{L}$. For a more sophisticated explanation how speckle noise appears in SAR we refer to [21].
In our regularization terms we use the following matrices D:

i) For the discrete TV functional we use

$$D := \begin{pmatrix} I \otimes D_0 \\ D_0 \otimes I \end{pmatrix}, \quad D_0 := \begin{pmatrix} -1 & 1 & & & \\ 0 & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & & 0 \end{pmatrix}, \quad (26)$$

where $\otimes$ denotes the tensor product (Kronecker product) of matrices. In our 1D computations, D is just the matrix D_0 without its last row.

ii) For the discrete NL-means functional we apply the following construction: For every image pixel $i = (i_1, i_2)$, we compute for all $j = (j_1, j_2)$ within a search window of size ω around i the distances

$$d_a(i,j) := \sum_{t_1=-p}^{p}\sum_{t_2=-p}^{p} G_a(t_1, t_2)\big(f(i_1 + t_1, i_2 + t_2) - f(j_1 + t_1, j_2 + t_2)\big)^2, \quad (27)$$

where G_a denotes the discretized, normalized Gaussian with standard deviation a. We refer to p as chosen patch size. Then we select the $\tilde{m}$ so-called 'neighbors' j of

i for which $d_a(i,j)$ takes the smallest values and set $w(i,j) = w(j,i) := 1$. Now, by (4) and (11), the matrix $D \in \mathbb{R}^{mn,n}$ with $m \leq 2\tilde{m}$ consists of m blocks of size $n \times n$, each having -1 as diagonal elements plus one additional nonzero value 1 in each row whose position is determined by the nonzero weights $w(i,j)$ and maybe some zero rows.

Our DRS algorithm reads

Algorithm I
Initialization: $u^{(0)} := f$, $b^{(0)} := 0$.
For $k = 0, 1, \ldots$ repeat until a stopping criterion is reached

$\qquad d^{(k+1)} := T_{\gamma\lambda}(b^{(k)} + Du^{(k)})$

$\qquad b^{(k+1)} := b^{(k)} + Du^{(k)} - d^{(k+1)}$

$\qquad u^{(k+1)}$ solution of

$\qquad\quad$ *Initialization:* $w^{(0)} := u^{(k)}$

$\qquad\quad$ For $j = 0, 1, \ldots$ solve until a stopping criterion is reached

$$\big((\tau + \gamma)I + \Lambda_{w^{(j)}} D^*D\big) w^{(j+1)} := \gamma f + w^{(j)}\big(D^*(d^{(k+1)} - b^{(k+1)}) + \tau \mathbb{1}\big).$$

Note that we have reversed the order of d, b and u in (15) - (17), where we would get the same iterates if we start with $d^{(0)} := Df$, $b^{(0)} := 0$. In our examples, it was possible to obtain positive vectors $u^{(k)}$ by choosing $\tau = 0$. Moreover, the solutions of the sparse linear systems can be realized by SOR iterations.

In [36], an algorithm based on the I-divergence - TV model was proposed for deblurring of images corrupted by multiplicative Poisson noise. Note that there doesn't exist a convergence proof here. Of course, we can also apply this algorithm without the blur operator for denoising tasks. Then the algorithm reads

Algorithm II
Initialization: $u^{(0)} := f$.
For $k = 0, 1, \ldots$ repeat until a stopping criterion is reached

$\qquad u^{(k+1)}$ solution of

$$\underset{v}{\operatorname{argmin}} \frac{1}{2}\big\|\frac{f - v}{\sqrt{u^{(k)}}}\big\|_2^2 + \lambda\Phi(Dv).$$

In [36], the authors propose to compute the minimizer by Chambolle's algorithm [7]. In the following computations, we use the gradient descent reprojection algorithm applied to the dual functional as suggested in [8]. Of course, one could also apply an DRS algorithm again.

We start with the 1D example depicted in Fig. 2. The figure and the noise level were chosen in accordance with the experiments in [39]. The right image shows the denoising

Algorithm	number of iterations		computational time
	inner	outer	
I	1	206	0.27 sec
II	7630	5	2.04 sec

Figure 1: Number of iterations and computational time of Algorithms I ($\gamma = 10$) and II for obtaining the result in Figure 2 up to a numerical error of 0.01.

result with the I-divergence - TV model, where the Algorithms I and II give the same result.

To compare the computational time needed, we ran both algorithms on a dual core desktop with two 2.4 GHz processors and 2.85 GB physical memory. For the results see Table 1. In this example, our Algorithm I with the 'backslash' MATLAB operation to solve the linear systems of equations is much faster than Algorithm II. To get a high accuracy in Algorithm II we have to perform a large number of inner iterations in the gradient descent reprojection algorithm.
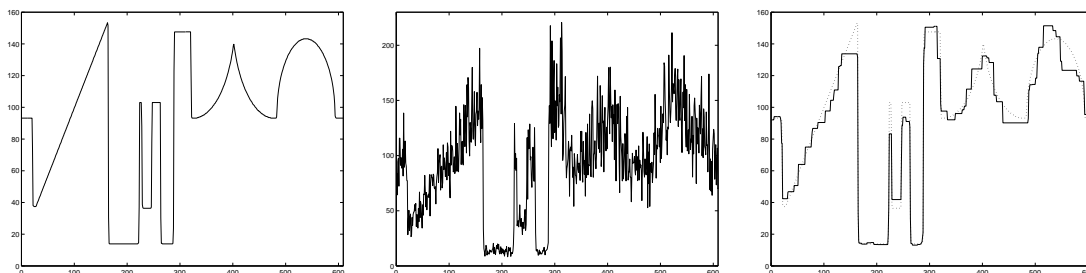

Figure 2: *Left*: Original signal. *Middle*: Noisy signal with Gamma noise of standard deviation 0.2. *Right*: Denoised signal by the I-divergence - TV model with $\lambda = 0.52$ (SNR= 15.86).

Next we compare the I-divergence - TV model with the I-divergence - NL-means model. Fig. 3 shows the original 'Barbara image' and its noisy version with Gamma noise of standard deviation 0.2. We compare the signal-to-noise ratios

$$SNR(u, g) := 20 \log_{10} \frac{\|g - \bar{g}\|_2}{\|(g - u) - \overline{(g - u)}\|_2},$$

where g denotes the original signal with mean $\bar{g}$ and u the denoised signal.

The denoising result obtained by the I-divergence - TV model are presented in Fig. 4. Again, the figure and the noise level were chosen to keep the experiments comparable with the ones in [39]. The images at the top show the result with the best SNR, while the image at the bottom has a better visual quality, of course with the usual staircasing. Note that

Figure 3: *Left*: Original image. *Right*: Noisy image corrupted with Gamma noise of standard deviation 0.2.

the loss of contrast responsible for the lower SNR was compensated by the affine rescaling of the `imagesc` MATLAB routine.

The results computed by the I-divergence - NL-means model are presented in Fig. 5. We have applied the I-divergence - NL-means model with $a = 3$, $p = 9$, $\omega = 11$ and $\tilde{m} = 5$. In the first iteration circle, we have computed D with respect to the logarithm of the noisy image $\log f$. If we use instead f for the computation of D, the resulting SNR becomes slightly worse. The image at the top shows the result after one iteration with respect to the best SNR. The image in the middle, which was also obtained after one iteration, was chosen with respect to the best visual quality after the affine rescaling. The image at the bottom of Fig. 5 shows the denoising result if we apply the algorithm a second time with the NL-means matrix D computed with respect to the denoised image of the first iteration circle (middle image). For f we have used again the original noisy image. As expected, the results by the I-divergence - NL-means model are significantly better than the results obtained by the I-divergence - TV model due to the semi-local adaptivity of the NL-means matrix D. Note that applying other than binary weights does not improve the denoising result.

Finally, Fig. 6 shows another denoising example with a more challenging noise level of standard deviation 0.5 , i.e., $L = 4$. Again, the I-divergence - NL-means model provides a very good reconstruction of the original image.

5   Conclusions

We have examined an I-divergence - TV model and an I-divergence - NL-means model for removing multiplicative noise, where we focused our attention in the numerical examples on Gamma noise. To the best of our knowledge this is the first paper which uses NL-means in the context of multiplicative noise and our numerical results are promising.
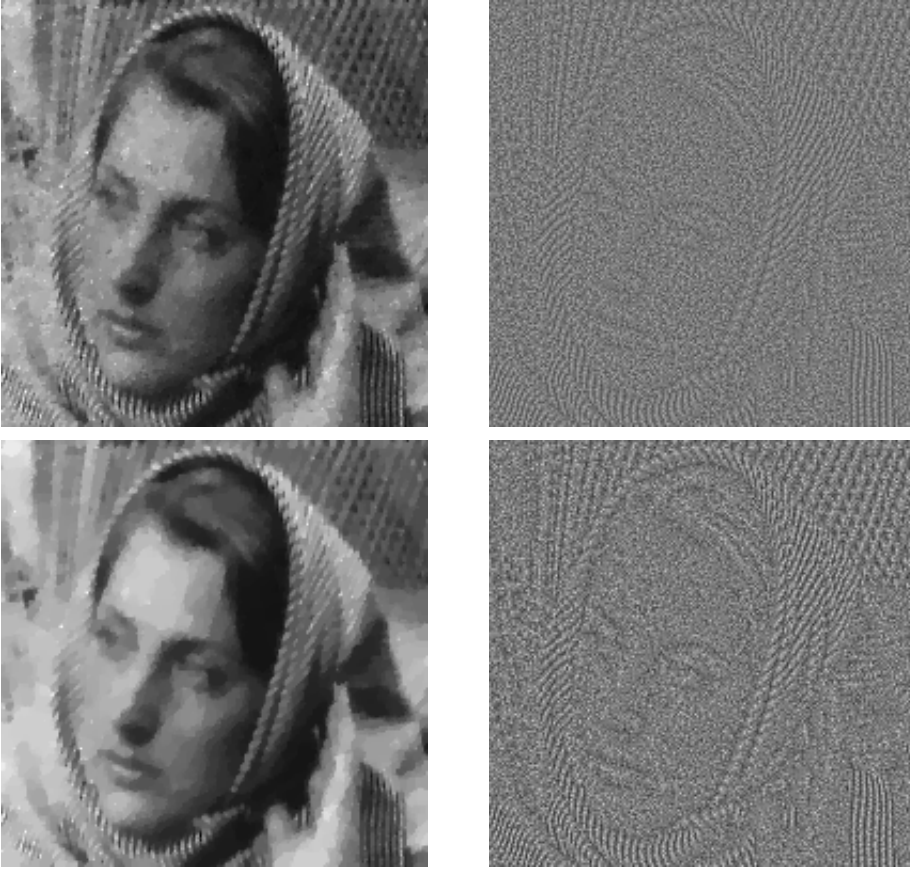
Figure 4: Denoised image (left) and quotient between the noisy image and the denoised image (right) by the I-divergence - TV model with $\lambda = 0.12$ (SNR = 11.06) (top) and with $\lambda = 0.2$ (SNR = 9.84) (bottom).

We proposed a Douglas-Rachford splitting technique (= alternating split Bregman algorithm) to find the minimizers of our functionals. The convergence of the Douglas-Rachford splitting algorithm is well-examined in the literature. However, one of our Bregman iteration step requires the solution of a nonlinear system of equations. We proposed a semi-implicit iterative algorithm to find this solution and proved the Q-linear convergence of the scheme.

Our approach can be simply generalized for restoring blurred images corrupted by, e.g., multiplicative Poisson noise. As a possible application we have electronic microscopy [27] in mind. More precisely, we intend to deblur the image by finding the minimizer of

$$\min_{u \in \mathbb{R}^n, u > 0} \left\{ \langle 1, Ku - f \log(Ku) \rangle + \lambda \Phi(Du) \right\},$$

where K with $K^*1 = 1$ denotes a blur matrix. Following (15) - (17), only the computation

of $u^{(k+1)}$ changes in our DRS algorithm; it is the solution of

$$\underset{u \in \mathbb{R}^n, u > 0}{\operatorname{argmin}} \langle 1, Ku - f \log(Ku) \rangle + \frac{1}{2\gamma} \|b^{(k)} + Du - d^{(k)}\|_2^2,$$

now, i.e., by setting the gradient to zero, the solution of the nonlinear system

$$0 = u - u K^* \frac{f}{Ku} + \frac{1}{\gamma} u \left(D^* Du + D^* (b^{(k)} - d^{(k)}) \right).$$

For the latter, one could use the semi-implicit algorithm

$$(\gamma I + \Lambda_{w^{(j)}} D^* D) w^{(j+1)} = w^{(j)} \left(\gamma K^* \frac{f}{K w^{(j)}} - D^* (b^{(k)} - d^{(k)}) \right). \tag{28}$$

However, the deblurring topic is beyond the scope of this paper; in particular we have to give a convergence proof of (28) in a forthcoming work.

References

[1] A. Aja-Fernández and C. Alberola-López. On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering. *IEEE Trans. on Imag. Proc.*, 15/9:2694–2701, 2005.

[2] F. Andreu-Vaillo, V. Caselles, and J. M. Mazón. *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals*. Birkhäuser Verlag, Basel - Boston - Berlin, 2004.

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

[4] J.-P. Aubin. *Optima and Equilibria: An Introduction to Nonlinear Analysis*. Springer, Berlin - Heidelberg - New York, 2nd edition, 1998.

[5] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

[6] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *IEEE Int. Conf. on Computer Vision and Pattern Recignition, CVPR*, volume 2, pages 60–65, 2005.

[7] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, (20):89–97, 2004.

[8] A. Chambolle. Total variation minimization and a class of binary MRF models. In A. Rangarajan, B. C. Vemuri, and A. L. Yuille, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition, EMMCVPR*, volume 3757 of *LNCS*, pages 136–152. Springer, 2005.

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

[10] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5–6):475–504, 2004.

[11] I. Csiszár. Why least squares and maximum entropy? an automatic approach to inference for linear inverse problems. *The Annals od Statistics*, 19/4:2032–2066, 1991.

[12] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Americ. Math. Soc.*, 82(2):421–439, 1956.

[13] S. Durand, J. Fadili, and M. Nikolova. Multiplicative noise cleaning via a variational method involving curvelet coefficients. In A. Lie, M. Lysaker, K. Morken, and X.-C. Tai, editors, *Scale Space and Variational Methods, Voss, Norway*. Springer, 2009.

[14] V. Dutt and J. F. Greenleaf. Adaptive speckle reduction filter for log-compressed B-scan images. *IEEE Trans. Med. Imaging*, 15/6:802–813, 1996.

[15] J. Eckstein and D. B. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[16] K. Frick. The Augmented Lagrangian Method and Associated Evolution Equations. 2008. Dissertation, University of Innsbruck.

[17] G. Gilboa, J. Darbon, S. Osher, and T. Chan. Nonlocal convex functionals for image regularization. *UCLA CAM Report*, 06-57, 2006.

[18] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Multiscale Modelling and Simulation*, 6(2):595–630, 2007.

[19] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *UCLA CAM Report*, 2007.

[20] D. Goldstein and S. Osher. The Split Bregman method for l_1 regularized problems. UCLA CAM Report, 2008.

[21] J. W. Goodman. Statistical properties of laser speckle patterns. In J. C. Dainty, editor, *Laser Speckle and Related Phenomena*, volume 9 of *Topics in Applied Physics*, pages 9–75. Springer, 1975.

[22] Y.-M. Huang, M. K. Ng, and Y.-W. Wen. A new total variation method for multiplicative noise removal. *Preprint*, 2008.

[23] M. F. Isana, R. F. Wagner, B. S. Garra, D. G. Brown, and T. H. Shawker. Analysis of ultrasound image texture via generalized rician statistics. *Opt. Eng.*, 25/6:743–748, 1986.

[24] E. Jakeman and P. N. Pusey. A model for non-rayleigh sea echo. *IEEE Trans. Antennas Propagat.*, 24/6:806–814, 1976.

[25] E. Jakeman and R. J. A. Tough. Generalized k distribution: a statistical model for weak scattering. *J. Opt. Soc. Amer.*, 4:1764–1772, 1987.

[26] K. Krissian, C.-F. Westin, and R. Kikinis. Oriented speckle reducing anisotropic diffusion. *IEEE Trans. on Imag. Proc.*, 2007.

[27] A. Kryvanos, J. Hesser, and G. Steidl. Nonlinear image restoration methods for marker extraction in 3d fluorescent microscopy. In *SPIE's 17th Annual Symposium EI05 - Electronic Imaging*. Springer, 2005.

[28] D. T. Kuan, A. A. Sawchuk, and T. C. Strand. Adaptive noise smoothing filter with signal dependent noise. *IEEE Trans. PAMI*, 7/2:165–177, 1985.

[29] J. Lee. Digital image enhancement and noise filtering using local statistics. *IEEE Trans. PAMI*, 2/2:165–168, 1980.

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

[31] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction.* SIAM, Philadelphis, 2001.

[32] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm. *IEEE Trans. on Nuclear Science*, 46/6:2202–2210, 1999.

[33] E. Resmerita, H. W. Engl, and A. N. Iusem. The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis. *Inverse Problems*, 23:2575–2588, 2007.

[34] L. I. Rudin, P.-L. Lions, and S. Osher. Multiplicative denoising and deblurring: Theory and algorithms. In S. Osher and N. Paragios, editors, *Geometric Level Sets in Imaging, Vision, and Graphics*, pages 103–120. Springer, 2003.

[35] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica A*, 60:259–268, 1992.

[36] A. Sawatzky, C. Brune, F. Wübbeling, T. Kösters, K. Schäfers, and M. Burger. Accurate EM-TV algorithm in PET with low SNR. *Preprint, University of Münster*, 2008.

[37] S. Setzer. Operator splittings, Bregman methods and and frame shrinkage in image processing. *Preprint, University of Mannheim*, 2009.

[38] S. Setzer. Split Bregman algorithm, Douglas-Racheford splitting and frame shrinkage. In A. Lie, M. Lysaker, K. Morken, and X.-C. Tai, editors, *Scale Space and Variational Methods*, LNCS, Voss, Norway, 2009.

[39] J. Shi and S. Osher. A nonlinear inverse scale space method for a convex multiplicative noise model. *SIAM Journal on Imaging Sciences*, 1(3):294–321, 2008.

[40] G. Steidl. A note on the dual treatment of higher order regularization functionals. *Computing*, 76:135–148, 2006.

[41] R. F. Wagner, S. W. Smith, and J. M. Sandrik. Statistics of speckle in ultrasound B-scans. *IEEE Trans. on Sonics Ultrason.*, 30/3:156–163, 1983.

[42] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[43] M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: a link between diffusion filtering and wavelet shrinkage. *Applied and Computational Harmonic Analysis*, 24:195–224, 2008.

[44] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168, 2008.

[45] Y. Yu and S. Acton. Speckle reducing anisotropic diffusion. *IEEE Trans. on Imag. Proc.*, 11/11:1260–1270, 2002.

[46] D. Zhou and B. Schölkopf. Regularization on discrete spaces. In *Pattern Recignition, Proc. of the 27th DAGM Symposium*, volume 2, pages 361–368, Berlin, Germany, 2005.
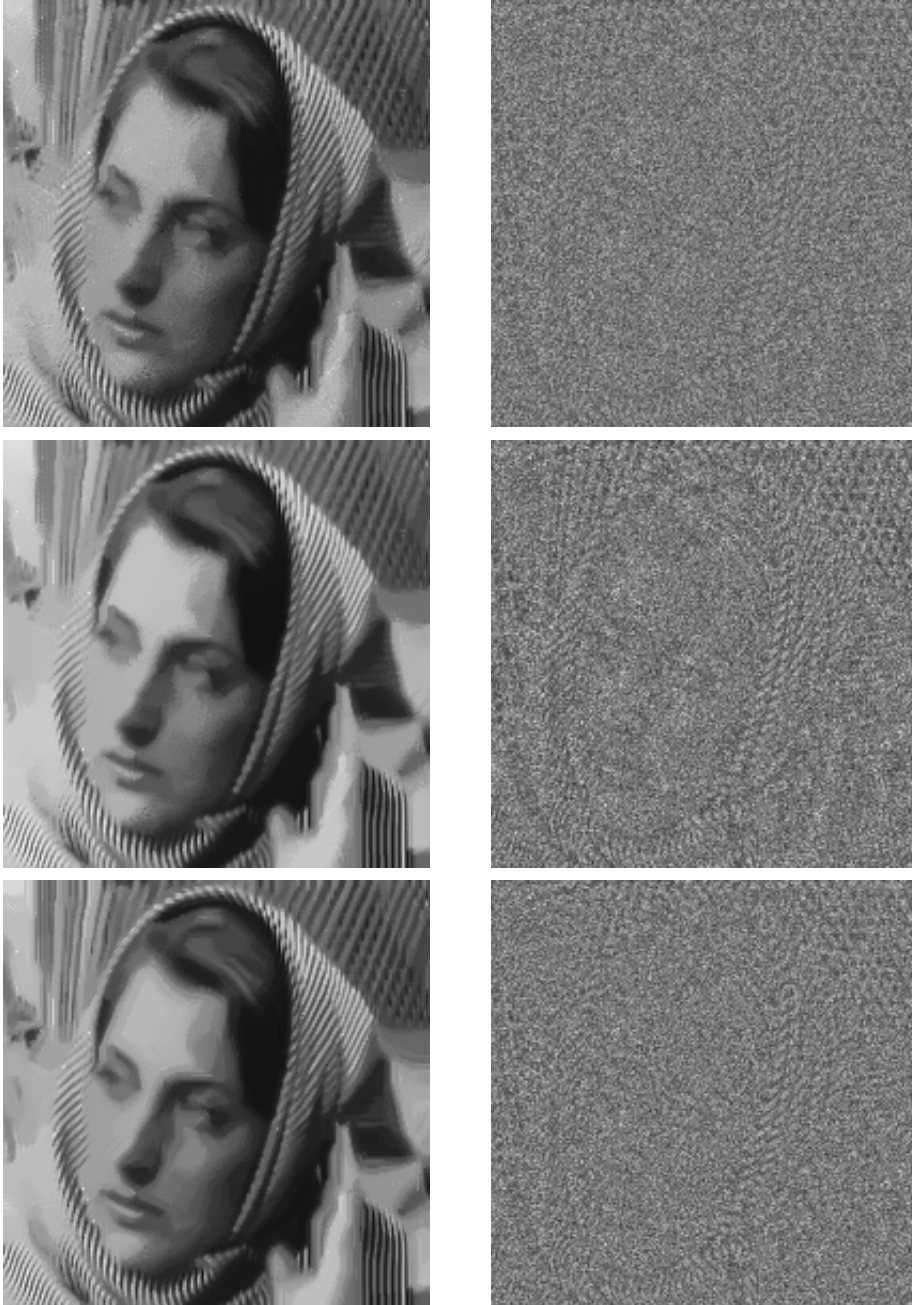
Figure 5: Denoised image (left) and quotient between the noisy image and the denoised image (right) by the I-divergence - NL-means model. *Top*: one iteration $\lambda = 0.08$ (SNR = 12.86). *Middle*: one iteration $\lambda = 0.12$ (SNR =11.71). *Bottom*: two iterations $\lambda = 0.09$ (SNR =12.82).

Figure 6: *Top:* Original 'cameraman image' (left) and corrupted image by multiplicative Gamma noise of standard deviation 0.5 (right). *Bottom:* Restored image by the I-divergence - TV model with $\lambda = 0.9$ (left) and the result by one iteration of the I-divergence - NL-means model with $\lambda = 0.6$, $a = 1.5$, $p = 5$, $\omega = 21$ and $\tilde{m} = 5$ (right).