Removing multiplicative noise by Douglas-Rachford splitting methods

G. Steidl and T. Teuber *

December 2, 2009

Abstract

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 for removing multiplicative Gamma noise. 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 minimizers of our restoration functionals by applying Douglas-Rachford splitting techniques, resp. alternating direction methods of multipliers. For a particular splitting, we present a semi-implicit scheme to solve the involved nonlinear systems of equations and prove its Q-linear convergence. Finally, we demonstrate the performance of our methods by numerical examples.

1 Introduction

We are interested in restoring images $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^2$ connected and bounded with Lipschitz boundary, arising from original images u, which are corrupted by (uncorrelated) multiplicative noise η of mean 1, i.e.,

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

The task of removing multiplicative noise appears in many applications, in particular in synthetic aperture radar (SAR) [9]. Here, we are confronted with speckle noise [40], which is usually assumed to follow a Gamma distribution. In electronic microscopy [45], single particle emission computed tomography (SPECT) [51] and positron emission tomography (PET) [56], non-additive Poisson noise appears in connection with blur.

In this paper, we focus on Gamma distributed noise although our model is appropriate for Poisson noise as well. Recently, various variational models for removing Gamma noise were proposed. Following the MAP estimator for multiplicative Gamma noise, Aubert and Aujol [4] introduced a non-convex model whose data term was subsequently adopted in a convex model by Shi and Osher in [59]. Indeed, these authors considered a more general data fitting term, which includes also the model in [54]. They applied a corresponding relaxed inverse scale space flow as denoising technique. The model of Shi and Osher was modified in [42] 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 [24].

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

Beyond variational approaches there exist other techniques to remove multiplicative noise, e.g., local linear minimum mean square error approaches [47, 46] or anisotropic diffusion methods [66, 1, 44] 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 [20] is usually applied as data fitting term. For the expectation-maximization (EM) approach related to the I-divergence model in deblurring problems see [49, 52] and the references therein and for the EM - total variation (TV) model we refer to [51, 56]. NL-means filters for removing non-additive noise were examined in [19, 43].
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 multiplicative 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 [17]. Note that for our setting the Douglas-Rachford splitting is equivalent to the alternating direction methods of multipliers (ADMM), also known as alternating split Bregman algorithm.
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 show that the minimizer of the functional proposed in [59] coincides with the minimizer of our I-divergence - TV model. Further properties of our restoration models for a discrete setting are proved in Section 3. In Section 4, we propose to compute the minimizers of our functionals by applying Douglas-Rachford splittings. For a particular splitting, we present a semi-implicit scheme to solve the involved nonlinear systems of equations and prove its Q-linear convergence. 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 5 and give a conclusion in Section 6.

2 Edge preserving variational methods for removing multiplicative noise

2.1 Review of models

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* depending on the given (corrupted) data f 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) [55]

$$|u|_{BV} := \sup_{p \in C_0^1, \| \, |p| \, \|_\infty \leq 1} \int_\Omega u \operatorname{div} p \, dx$$

which reads for $L_1(\Omega)$ functions with weak first derivatives in $L_1(\Omega)$ as

$$|u|_{BV} = \int_\Omega |\nabla u| \, dx. \tag{2}$$

The space $BV(\Omega)$ of *functions of bounded variation* consists of all $L_1(\Omega)$ functions with

$$\|u\|_{BV} := \|u\|_{L_1} + |u|_{BV} < \infty.$$

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

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

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 $f_{\min} \le \hat{u}_i \le f_{\max}$, 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 [12] and was incorporated into the variational framework in [35, 36, 37, 67]. We refer to these papers for further information on NL-means. Based on some pre-computed weights w the regularization term is given by

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

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 the functional to minimize 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 data 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

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

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 smaller 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:

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

AA-model. Based on the MAP estimator for multiplicative Gamma noise, Aubert and Aujol [4] proposed to determine the denoised image as a minimizer in $\{u \in BV : u > 0\}$ of the following, in general non-convex, functional

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

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 [4, 59] 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 $0 < f_2 \leq f_1 \leq 3f_2$. 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

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

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 [59] suggested to keep the data fitting term in (7) but to replace the regularizer $|u|_{BV}$ by $|\log u|_{BV}$. Moreover, setting $w := \log u$ as in the log-model, this results in the convex functional

$$\hat{w} = \operatorname*{argmin}_{w \in BV} \Big\{ \int_\Omega f e^{-w} + w \, dx + \lambda |w|_{BV} \Big\}, \quad \hat{u} = e^{\hat{w}} \tag{8}$$

which overcomes the drawback of (7) as we will see by looking at our two-pixel model (5). This model was also considered in paper [7] written in parallel to our manuscript.

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 $afe^{-w} + \frac{b}{2} f^2 e^{-2w} + (a+b)w$ which includes also the model in [54], 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 Poisson 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* [10] of the *Boltzmann-Shannon entropy.* 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

$$\hat{u} = \operatorname*{argmin}_{u \in BV, u > 0} \Big\{ \int_\Omega u - f \log u \, dx + \lambda |u|_{BV} \Big\}. \tag{9}$$

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.

2.2 Relation between the SO-model and the I-divergence model

In this subsection, we will see that the minimizers $\hat{w}$ and $\hat{u}$ of the functionals in the SO-model (8) and the I-divergence model (9), resp., coincide in the sense that $\hat{u} = e^{\hat{w}}$. Therefore, we prefer to work with the well-examined I-divergence model in the subsequent sections, which does not require to take finally the exponent of the minimizing function. Let us start with some rough arguments to see the basic idea.

Remark 2.1. *(An heuristic argument)*
Since $\nabla e^w = e^w \nabla w$ we have for $u = e^w$ that $\nabla u(x) = 0$ if and only if $\nabla w(x) = 0$. If we minimize over smooth functions, the minimizers $\hat{w}$ and $\hat{u}$ of (8) *and* (9)*, resp., are given by*

$$0 = 1 - fe^{-\hat{w}} - \lambda \operatorname{div} \frac{\nabla \hat{w}}{|\nabla \hat{w}|}, \quad 0 = 1 - \frac{f}{\hat{u}} - \lambda \operatorname{div} \frac{\nabla \hat{u}}{|\nabla \hat{u}|}$$

for $|\nabla \hat{w}(x)| \neq 0$ and $|\nabla \hat{u}(x)| \neq 0$. Since $\frac{\nabla w}{|\nabla w|} = \frac{e^w \nabla w}{e^w |\nabla w|} = \frac{\nabla u}{|\nabla u|}$, we obtain that indeed $\hat{u} = e^{\hat{w}}$. Obviously, this approach works also for anisotropic TV regularizers, e.g., the functional considered in [61]*.*

The above argumentation has two gaps. First, we have to use $|u|_{BV}$ instead of $\int |\nabla u| \, dx$ in BV. Second, the data fitting terms in (8) and (9) are not continuous over L_p, $1 \le p < \infty$ and over BV. The rest of this subsection closes these gaps.

For a function $f \in L_\infty(\Omega)$, let $f_{\min} := \operatorname{ess\,inf}_x f(x)$ and $f_{\max} := \operatorname{ess\,sup}_x f(x)$. In the following, we restrict our attention to functions $f \in L_\infty(\Omega)$ with $f_{\min} > 0$.
Consider the integrands $\varphi, \psi : \Omega \times \mathbb{R} \to [0, +\infty]$ defined by

$$\varphi(x,s) := s + f(x)e^{-s} - \log f(x) - 1, \tag{10}$$

$$\psi(x,s) := \begin{cases} s - f(x)\log s + f(x)\log f(x) - f(x) & \text{for } s > 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{11}$$

For integrands, properties like continuity, lower semi-continuity (l.s.c.), convexity and sub-differentiability are understood with respect to the second variable. For fixed $x \in \Omega$, the functions φ and ψ have their minimum 0 in $s = \log f(x)$ and $s = f(x)$, respectively. Fig. 1 shows the functions φ, ψ for $f(x) \in \{e^{-1}, 1, e\}$.

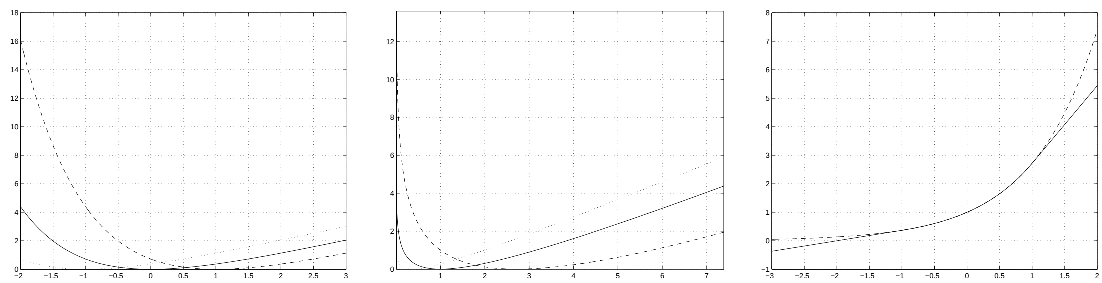

Figure 1: *Left:* The function $\varphi(x,s)$ for $f(x) = e^{-1}$ (dotted), $f(x) = 1$ (solid) and $f(x) = e$ (dashed). *Middle:* The function $\psi(x,s)$ for the same values of $f(x)$. *Right:* The function g_k for $a_k = -1$ and $b_k = 1$ (solid) and the exponential function (dashed).

The functionals $S_\varphi, S_\psi : L_1(\Omega) \to [0, +\infty]$ given by

$$S_\varphi(w) := \int_\Omega \varphi(x, w(x)) \, dx, \quad S_\psi(u) := \int_\Omega \psi(x, u(x)) \, dx$$

are proper. Since $\varphi(x, \cdot)$ and $\psi(x, \cdot)$ are normal and strictly convex, S_ϕ and S_ψ are strictly convex, too. Moreover, by [53] one has that

$$\begin{aligned} \partial S_\varphi(w) &= \{w^* \in L_\infty : w^*(x) \in \partial\varphi(x, w(x)) \ \text{ a.e. } x \in \Omega\} \\ &= \begin{cases} 1 - fe^{-w} & \text{for } 1 - fe^{-w} \in L_\infty, \\ \emptyset & \text{otherwise,} \end{cases} \\ \partial S_\psi(u) &= \{u^* \in L_\infty : u^*(x) \in \partial\psi(x, u(x)) \ \text{ a.e. } x \in \Omega\} \\ &= \begin{cases} 1 - f/u & \text{for } u > 0 \text{ a.e. and } 1 - f/u \in L_\infty, \\ \emptyset & \text{otherwise.} \end{cases} \end{aligned}$$

Hence, the equality

$$\partial S_\varphi(w) = \partial S_\psi(e^w)$$

holds true. Since φ and ψ are nonnegative, the functionals S_ϕ and S_ψ are l.s.c., see [26, p. 239]. Note that there exist $C_\phi, C_\psi > 0$ and $D_\phi, D_\psi > 0$ such that

$$\varphi(x,s) \geq C_\varphi |s| - D_\varphi \quad \text{and} \quad \psi(x,s) \geq C_\psi |s| - D_\psi. \tag{12}$$

Choose for example $a < \min\{0, \log(f_{\min})\}$ and $b > \max\{0, \log(f_{\max})\}$, set $C_a := 1 - f_{\min} e^{-a} < 0$, $C_b := 1 - f_{\max} e^{-b} > 0$ and $w_a := a + f_{\min} e^{-a} - \log(f_{\min}) - 1$, $w_b := b + f_{\max} e^{-b} - \log(f_{\max}) - 1$, $u_b := e^b - f_{\max} b + f_{\max} \log(f_{\max}) - f_{\max}$ and take $C_\varphi = \min\{|C_a|, C_b\}$, $C_\psi = C_b$ and $D_\varphi = \max\{C_a a - w_a, C_b b - w_b\}$, $D_\psi = C_b e^b - u_b$.
By properties of $|\cdot|_{BV}$, see, e.g., [29], we obtain that the functionals $T_\varphi, T_\psi : BV(\Omega) \to [0, +\infty]$ defined by

$$T_\varphi(w) := S_\varphi(w) + \lambda |w|_{BV}, \quad T_\psi(u) := S_\psi(u) + \lambda |u|_{BV} \tag{13}$$

are also proper, l.s.c. and strictly convex. By the following lemma, the functionals T_φ and T_ψ are BV-coercive such that by [29, p. 176] there exists a minimizer of both functionals which is unique due to their strict convexity.

Lemma 2.2. *Let $f \in L_\infty(\Omega)$ with $f_{\min} > 0$. Then T_φ and T_ψ are BV-coercive.*

Proof: Since the following arguments are the same for T_φ and T_ψ, we consider only the later one. By the Poincaré inequality [3, p. 302] there exists a constant C_P such that

$$\|u - \bar{u}\|_{L_1} \leq C_P |u|_{BV}, \quad \bar{u} := \frac{1}{\mu(\Omega)} \int_\Omega u(x)\, dx,$$

where μ denotes the Lebesgue measure on $\mathbb{R}^2$. We obtain further

$$\begin{aligned} \|u\|_{L_1} &\leq \|u - \bar{u}\|_{L_1} + |\bar{u}|\, \mu(\Omega) \leq C_P |u|_{BV} + |\bar{u}|\, \mu(\Omega), \\ \|u\|_{BV} &\leq (C_P + 1)|u|_{BV} + |\bar{u}|\, \mu(\Omega) \leq C\, T_\psi(u) + |\bar{u}|\, \mu(\Omega), \end{aligned}$$

where $C := (C_P + 1)/\lambda$. By (12), we see that

$$S_\psi(u) \geq C_\psi \int_\Omega |u(x)|\, dx - D_\psi\, \mu(\Omega) \geq \mu(\Omega)(C_\psi |\bar{u}| - D_\psi)$$

and consequently

$$|\bar{u}| \leq \frac{1}{\mu(\Omega) C_\psi} S_\psi(u) + \frac{D_\psi}{C_\psi} \leq \frac{1}{\mu(\Omega) C_\psi} T_\psi(u) + \frac{D_\psi}{C_\psi}.$$

Finally, we get

$$\|u\|_{BV} \leq (C + 1/C_\psi) T_\psi(u) + \mu(\Omega) D_\psi / C_\psi. \qquad \square$$

For another proof of the BV coercivity of T_ψ see [11]. We will need the facts from the following theorem.

Theorem 2.3. i) *Let $h : \mathbb{R} \to \mathbb{R}$ be a non-decreasing and Lipschitz continuous function. Let $u \in BV(\Omega)$. Then $h(u) \in BV$ and $\partial R(u) \subset \partial R(h(u))$.*
ii) *Let $\phi : \Omega \times \mathbb{R} \to (-\infty, +\infty]$ be a measurable function. Assume that there exists a nonnegative function $\gamma \in L_1(\Omega)$, a constant $C > 0$ and $1 \leq p < \infty$ such that*

$$|\phi(x,s)| \leq C|s|^p + \gamma(x) \quad \text{a.e. } x \in \Omega,\ \forall s \in \mathbb{R}. \tag{14}$$

Then the functional S_ϕ is L_p continuous if and only if $\phi(x, \cdot)$ (or equivalent integrands) is continuous a.e. $x \in \Omega$.

For the proof of i) see [57, p. 148] and for ii) see [31, p. 442]. For i) with locally absolutely continuous functions h see [41].
The final theorem of this subsection is a special instance of a more general theorem proved by M. Grasmair in [41]. We found it useful to give this simplified version for our special setting.

Theorem 2.4. *Let T_φ and T_ψ be given by (13), where $f \in L_\infty(\Omega)$ with $f_{\min} > 0$. Then $\hat{w}$ is the minimizer of T_φ if and only if $\hat{u} = e^{\hat{w}}$ is the minimizer of T_ψ.*

Proof: We show that $\hat{w} = \operatorname{argmin}_w T_\varphi(w)$ implies $e^{\hat{w}} = \operatorname{argmin}_u T_\psi(u)$. The reverse direction follows since the minimizers are unique.
The idea is to approximate the integrands φ, ψ and the exponential function by some 'nicer' functions for which the 'adapted' assertion follows immediately by Theorem 2.3 and to apply Γ-convergence arguments to get the final result.
1. To this end, choose a sequence of increasing intervals $[a_k, b_k] \subset [a_{k+1}, b_{k+1}]$ such that $\bigcup_k [a_k, b_k] = \mathbb{R}$ and $a_k < \min\{0, \log(f_{\min})\}$, $b_k > \max\{0, \log(f_{\max})\}$ for all $k \in \mathbb{N}$. Let

$$\xi_k(x) := 1 - f(x)e^{-a_k} < 0, \quad \zeta_k(x) := 1 - f(x)e^{-b_k} > 0$$

and $\xi_k := \operatorname{ess\,inf}_x \xi_k(x) > -\infty$, $\zeta_k := \operatorname{ess\,sup}_x \zeta_k(x) < \infty$. Now we define the truncated continuous integrands for a.e. $x \in \Omega$ by

$$\varphi_k(x,s) := \begin{cases} \varphi(x,a_k) + \xi_k(x)(s-a_k) & \text{if} \quad s < a_k, \\ \varphi(x,s) & \text{if} \quad s \in [a_k, b_k], \\ \varphi(x,b_k) + \zeta_k(x)(s-b_k) & \text{if} \quad s > b_k, \end{cases}$$

$$\psi_k(x,s) := \begin{cases} \psi(x,e^{a_k}) + \xi_k(x)(s-e^{a_k}) & \text{if} \quad s < e^{a_k}, \\ \psi(x,s) & \text{if} \quad s \in [e^{a_k}, e^{b_k}], \\ \psi(x,e^{b_k}) + \zeta_k(x)(s-e^{b_k}) & \text{if} \quad s > e^{b_k}. \end{cases}$$

Let

$$T_{\varphi,k}(w) := S_{\varphi_k}(w) + \lambda |w|_{BV}, \quad T_{\psi,k}(u) := S_{\psi_k}(u) + \lambda |u|_{BV}.$$

For any $k \in \mathbb{N}$, these functionals are proper, l.s.c., convex. Moreover, by the same arguments as in the proof of Lemma 2.2, we see that

$$T_{\varphi,k}(w) \geq C_1 \|w\|_{BV} - D_1, \quad T_{\psi,k}(u) \geq C_2 \|u\|_{BV} - D_2 \quad \forall k \in \mathbb{N} \tag{15}$$

for some $C_1, C_2, D_1, D_2 > 0$ so that the functionals are equi-coercive on BV. Recall that a sequence $\{T_k\}_k$ is *equi-coercive* on a metric space X if and only if there exists a l.s.c., coercive function $F : X \to (-\infty, +\infty]$ such that $T_k \geq F$ for all k, see [21, Proposition 7.7]. Hence, there exist minimizers $\hat{w}_k$ and $\hat{u}_k$ of $T_{\varphi,k}$ and $T_{\psi,k}$, respectively.
Since $f \in L_\infty$ with $f_{\min} > 0$, the functions φ_k fulfill condition (14) with $C_k := \max\{|\xi_k|, \zeta_k\}$, $p = 1$ and $\gamma_k := \varphi_k(\cdot, 0) \in L_1(\Omega)$. Therefore, we have by Theorem 2.3 ii) that S_{φ_k} is continuous on L_1 (and on BV). Set

$$g_k(s) := \begin{cases} e^{a_k} + e^{a_k}(s-a_k) & \text{if} \quad s < a_k, \\ e^s & \text{if} \quad s \in [a_k, b_k], \\ e^{b_k} + e^{b_k}(s-b_k) & \text{if} \quad s > b_k, \end{cases}$$

see Fig. 1 (right). This truncated exponential function is a non-decreasing Lipschitz continuous function, so that we get by 2.3 i) that $\partial |w|_{BV} \subset \partial |g_k(w)|_{BV}$ for all $w \in BV(\Omega)$.

Further, one can check by straightforward computation that $\partial\varphi_k(x,s) = \partial\psi_k(x, g_k(s))$ so that $\partial S_{\varphi_k}(w) = \partial S_{\psi_k}(g_k(w))$. Thus, we obtain by [26, p. 26] that

$$\partial T_{\varphi,k}(w) = \partial S_{\varphi_k}(w) + \lambda\partial|w|_{BV} \subset \partial S_{\psi_k}(g_k(w)) + \lambda\partial|g_k(w)|_{BV} \subset \partial T_{\psi,k}(g_k(w)).$$

Now $\hat{w}_k$ is a minimizer of $T_{\varphi,k}$ iff $0 \in \partial T_{\varphi,k}(\hat{w}_k)$. Hence, if $\hat{w}_k$ is a minimizer of $T_{\varphi,k}$, then $\hat{u}_k = g_k(\hat{w}_k)$ is a minimizer of $T_{\psi,k}$.
2. The sequences $\{\varphi_k(x, w(x))\}_k$ and $\{\psi_k(x, u(x))\}_k$ are increasing sequences of nonnegative functions and

$$\lim_{k\to\infty} \varphi_k(x, w(x)) = \varphi(x, w(x)), \quad \lim_{k\to\infty} \psi_k(x, u(x)) = \psi(x, u(x)) \quad \text{a.e. } x \in \Omega,$$

so that by the Monotone Convergence Theorem

$$\lim_{k\to\infty} S_{\varphi_k} = S_\varphi, \quad \lim_{k\to\infty} S_{\psi_k} = S_\psi.$$

Since $\{\varphi_k(x, w(x))\}_k$ and $\{\psi_k(x, u(x))\}_k$ are increasing we see that $\{T_{\varphi,k}\}_k$ and $\{T_{\psi,k}\}_k$ are increasing sequences, too. Therefore we have by [8, p. 35] that

$$\Gamma-\lim_{k\to\infty} T_{\varphi,k} = \lim_{k\to\infty} T_{\varphi,k} = T_\varphi, \quad \Gamma-\lim_{k\to\infty} T_{\psi,k} = \lim_{k\to\infty} T_{\psi,k} = T_\psi. \tag{16}$$

3. By the Γ-convergence (16) and since the functionals $T_{\varphi,k}$ and $T_{\psi,k}$ are equi-coercive it follows that $T_{\varphi,k}(\hat{w}_k) \to T_\varphi(\hat{w})$ and $T_{\psi,k}(g_k(\hat{w}_k)) \to T_\varphi(\hat{u})$ and that each cluster point of $\{\hat{w}_k\}_k$ is $\hat{w}$ and of $\{g_k(\hat{w}_k)\}_k$ is $\hat{u}$, see [8, p. 29].
4. Let $\{\hat{w}_n\}_{n\in I}$, $\{g_n(\hat{w}_n)\}_{n\in I}$, $I \subset \mathbb{N}$, be subsequences converging in L_1 to $\hat{w}$ and $\hat{u}$, resp. By [2, Theorem 13.6], there exists a subsequence $\{\hat{w}_m\}_{m\in I_1}$, $I_1 \subset I$ which converges a.e. to $\hat{w}$. Then by construction of g_m, we have that $g_m(\hat{w}_m)$ converges a.e. to $e^{\hat{w}}$. On the other hand, we know that $g_m(\hat{w}_m)$ converges in L_1 to $\hat{u}$. Hence, there is a subsequence $\{g_k(\hat{w}_k)\}_{k\in I_2}$, $I_2 \subset I_1$ which converges a.e. to $\hat{u}$ and consequently $\hat{u} = e^{\hat{w}}$ a.e. □

3 Discrete denoising model

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

i) some discretization of the gradient operator as, e.g., those in [14, 60] with $m = 2$, see (36), or

ii) the NL-means operator with binary weights introduced in [37] with m associated to the number of permitted neighbors, see Section 5. 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

$$\min_{u\in\mathbb{R}^n} \big\{\Psi(u) + \lambda \underbrace{\phi(Du)}_{\Phi(u)}\big\}, \tag{17}$$

with

$$\Psi(u) := \begin{cases} \langle 1, u - f\log u \rangle & \text{if } \; u > 0, \\ \infty & \text{otherwise} \end{cases} \quad \text{and} \quad \phi(p) := \|\,|p|\,\|_1. \tag{18}$$

If D is given by i) then we refer to (17) as I-divergence - TV model and if D is determined by ii) we call it I-divergence - NL-means model. The functional in (17) is proper, l.s.c., coercive and strictly convex on its domain. Therefore, there exists a unique minimizer. The dual problem of (17) reads

$$-\min_{p \in \mathbb{R}^{mn}} \left\{ \Psi^*(-D^*p) + \lambda \phi^*(\lambda^{-1}p) \right\} \tag{19}$$

with the conjugate functions

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

where

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

denotes the *indicator function* of $C := \{p \in \mathbb{R}^{mn} : \|\,|p|\,\|_\infty \le 1\}$. There is no duality gap, i.e., (17) and (19) take the same value and the minimizers are related by

$$\hat{u} = f/(1 + D^*\hat{p}). \tag{20}$$

The following proposition describes properties of the minimizer of (17).

Proposition 3.1. *The solution* $\hat{u}$ *of* (17) *has the following properties:*

i) Minimum-maximum principle:

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

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 [4, Theorem 4.1]. We have only to verify the relations

$$\Phi\big(\min(u, f_{\max})\big) \le \Phi(u), \quad \Phi\big(\max(u, f_{\min})\big) \le \Phi(u).$$

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

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

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 (20) and since $1^{\mathrm{T}} D^* = 0$. Namely, we have

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

The following proposition deals with the SO-model versus our I-divergence model in the discrete setting.

Proposition 3.2. *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 (17) coincides with $e^{\hat{w}}$, where $\hat{w}$ is the minimizer of the discrete SO-model*

$$\hat{w} := \operatorname*{argmin}_{w \in \mathbb{R}^n} \{\langle 1, f e^{-w} + w \rangle + \lambda \phi(Dw)\}. \tag{21}$$

Proof. We have that $\hat{u}$ and $\hat{w}$ are the minimizers of (17) and (21), 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., [3] 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

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

The equality can be rewritten as

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

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

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

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

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

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

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, if the $(i+jn)$–th row of D contains 1 at column $(i+jn)_1$ and -1 at column $(i+jn)_2$, we get that

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

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

Note that the proof shows that the discrete models are also identical for $m \geq 2$ in the anisotropic setting $\phi(p) := \|p\|_1$.

4 Minimization by Douglas-Rachford splitting

4.1 Douglas-Rachford splitting and ADMM

We consider problems of the form

$$\min_{u \in \mathbb{R}^n} \{\Psi(u) + \lambda \underbrace{\phi(Du)}_{\Phi(u)}\}, \tag{22}$$

where $\Psi : \mathbb{R}^n \to (\infty, +\infty]$ and $\phi : \mathbb{R}^{mn} \to (\infty, +\infty]$ are proper, l.s.c., convex and $D \in \mathbb{R}^{mn,n}$, as well as their dual problems

$$-\min_{p \in \mathbb{R}^{mn}} \{\Psi^*(-D^* p) + \lambda \phi^*(\lambda^{-1} p)\}. \tag{23}$$

By Fermat's rule we know that $\hat{p}$ is a minimizer of (23) if and only if

$$0 \in \partial\left(\Psi^* \circ (-D^*)\right)(\hat{p}) + \partial\phi^*(\lambda^{-1}\hat{p}). \tag{24}$$

Since Ψ^* and ϕ^* are proper, l.s.c. and convex, the (in general) set-valued operators $\partial(\Psi^* \circ (-D^*))$ and $\partial\phi^*$ are maximal monotone, see [5]. 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 [5]. Inclusions of the form (24) can be solved by various splitting techniques like *forward-backward splitting* or *Douglas-Rachford splitting* (DRS), see [25, 48]. In this paper, we focus on the DRS because it leads to an efficient algorithm. Note that DRS was first considered in [23] for linear operators.

Theorem 4.1. *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 $\big(p^{(k)}\big)_{k\in\mathbb{N}}$ converges to $\hat{p}$.

For the proof see, e.g., [16]. Note that the iterates in the DRS need not to be computed exactly. Their convergence is also ensured if we allow summable errors, cf., [25].

The DRS applied to the dual problem (24) is related to the ADMM. To introduce this algorithm we consider the equivalent problem of (17)

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

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

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

If we replace $b^{(k)}$ by $b^{(k)}/\gamma$, this method is also known in image processing as *split Bregman algorithm*, see [13, 32, 39, 62, 65]. Since the first functional is in general hard to minimize, one uses instead the following *alternating split Bregman algorithm* (25) - (27). As pointed out by Esser in [27], this algorithm can be traced back to the ADMM proposed in [34, 38]. Therefore, we will refer to this algorithm as ADMM:

Algorithm

$$u^{(k+1)} = \operatorname*{argmin}_{u\in\mathbb{R}^n} \big\{\Psi(u) + \frac{1}{2\gamma}\|b^{(k)} + Du - d^{(k)}\|_2^2\big\}, \tag{25}$$

$$d^{(k+1)} = \operatorname*{argmin}_{d\in\mathbb{R}^{mn}} \big\{\phi(d) + \frac{1}{2\gamma\lambda}\|b^{(k)} + Du^{(k+1)} - d\|_2^2\big\}, \tag{26}$$

$$b^{(k+1)} = b^{(k)} + Du^{(k+1)} - d^{(k+1)}. \tag{27}$$

The convergence of $d^{(k)}$ and $b^{(k)}$ is ensured by Theorem 4.1 and the following proposition from [25, 33], see also [58].

Proposition 4.2. *The above ADMM 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*

$$t^{(k)} = \eta(b^{(k)} + d^{(k)}), \quad p^{(k)} = \eta b^{(k)}, \quad k \geq 0. \tag{28}$$

Moreover, the convergence of $\{u^{(k)}\}_{k\in\mathbb{N}}$ defined in (25) to a solution of the primal problem is guaranteed if the primal problem has a unique minimizer or if $\operatorname{argmin}_{u\in\mathbb{R}^n}\{\Psi(u) + \frac{1}{2\gamma}\|\hat{b} + Du - \hat{d}\|_2^2\}$ has a unique solution.

4.2 ADMM for the I-divergence - TV/NL-means model

First, we apply the ADMM to our setting in (18). By the simple structure of ϕ in (18), the solution of (26) 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 $p = (\mathrm{p}_1, \dots, \mathrm{p}_m)^{\mathrm{T}} \in \mathbb{R}^{mn}$ with $\mathrm{p}_k := (p_{j+(k-1)n})_{j=1}^n$ by

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

see, e.g., [15, 64].
In contrast to the original problem (17), the functional (25) in the ADMM has a quadratic penalizer which is differentiable. Setting the gradient to zero, we see that w is a solution of (25) if $w > 0$ satisfies

$$0 = 1 - \frac{f}{w} + \frac{1}{\gamma}\left(D^*Dw + D^*(b^{(k)} - d^{(k)})\right), \tag{29}$$

$$0 = \gamma w - \gamma f + w\, D^*Dw - w\, D^*(d^{(k)} - b^{(k)}). \tag{30}$$

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 [17].

Remark 4.3. *We apply DRS in* (29) *with*

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

and obtain

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

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 (30) by an efficient method which can be deduced directly from (30) by adding $\tau\, w$ for some $\tau \ge 0$ 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 $j = 0, 1, \dots$ 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 (30). 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 $D = (d_1 \dots d_n)$ 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 4.4. *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*

$$\big((\tau + \gamma)I + \Lambda_{w^{(j)}} A\big)\, w^{(j+1)} = \gamma f + w^{(j)}\, (c + \tau\, 1) \tag{31}$$

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

$$0 = \gamma w - \gamma f + w\, Aw - w\, c. \tag{32}$$

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 the right-hand side of (31) 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 (31) 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 (31) and (32) 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 (32) that $Aw - c = \gamma(f/w - 1)$ so that

$$\begin{aligned} \big((\tau+\gamma)I + A_j\big)(w - w^{(j+1)}) &= (\gamma(f/w - 1) - \tau\, 1)\,(w^{(j)} - w), \\ \big(I + \frac{1}{\tau+\gamma} A_j\big)(w - w^{(j+1)}) &= \frac{\tau\, 1 - \gamma(f/w - 1)}{\tau + \gamma}\,(w - w^{(j)}). \end{aligned} \tag{33}$$

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

$$\frac{\|\tau\, 1 - \gamma(f/w - 1)\|_\infty}{\tau + \gamma} < 1. \tag{34}$$

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 (33) and (34) this implies

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

so that the algorithm converges Q-linearly. □

Remark 4.5. *The proof of Theorem 4.4 poses two conditions on τ. In our practical applications these conditions can be relaxed to $\tau = 0$ in most cases by the following reasons: If $f/w - 1 \in (-1, 1)$, i.e., $0 < f/w < 2$, then $\tau = 0 > \gamma\, (\frac{1}{2}\, \|f/w\|_\infty - 1)$. Indeed, this setting is realistic, since f is approximately a noisy variant of w, i.e., $(1-\epsilon)w \leq f \leq (1+\epsilon)w$, $\epsilon \in (0,1)$. The first condition on τ that the right-hand side in (31) is positive may lead to large numbers τ so that the convergence factor $\alpha \approx 1$ and the convergence is very slow. However, it follows directly from the proof that a fast convergence is guaranteed as long as $w^{(j)}$ remains positive during the iteration process.*

In summary, we finally end up with the following algorithm:

Algorithm

Initialization: $d^{(0)} := Df$, $b^{(0)} := 0$.
For $k = 0, 1, \ldots$ repeat until a stopping criterion is reached
  Set $u^{(k+1)}$ to be the final iterate of
    Initialization: $w^{(0)} := u^{(k)}$
    For $j = 0, 1, \ldots$ solve until a stopping criterion is reached

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

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

Up to now we have considered the ADMM with respect to the splitting

$$\text{(I)} \qquad \Psi(u) + \lambda\phi(d) \quad \text{subject to} \quad d = Du,$$

which requires in

- Step 1: the solution of a nonlinear system of equations,

- Step 2: coupled shrinkage.

Other splittings are possible as

$$\begin{array}{lll} \text{(II)} & \Psi(u) + \lambda\Phi(v) & \text{subject to} \quad v = u. \\ \text{(III)} & \langle 0, v\rangle + \Psi(u) + \lambda\phi(d) & \text{subject to} \quad \begin{pmatrix} u \\ d \end{pmatrix} = \begin{pmatrix} I \\ D \end{pmatrix} v. \end{array}$$

The ADMM for (II) reads

$$\begin{aligned} u^{(k+1)} &= \operatorname*{argmin}_{u \in \mathbb{R}^n} \{\Psi(u) + \frac{1}{2\gamma}\|b^{(k)} + u - v^{(k)}\|_2^2\}, \\ v^{(k+1)} &= \operatorname*{argmin}_{v \in \mathbb{R}^{mn}} \{\Phi(v) + \frac{1}{2\gamma\lambda}\|b^{(k)} + u^{(k+1)} - v\|_2^2\}, \\ b^{(k+1)} &= b^{(k)} + u^{(k+1)} - v^{(k+1)} \end{aligned}$$

and can be tackled as follows:

- Step 1: This minimization problem can be decoupled and then solved componentwise by

$$u^{(k+1)} = \frac{1}{2}\left(v^{(k)} - b^{(k)} - \gamma + \sqrt{(v^{(k)} - b^{(k)} - \gamma)^2 + 4\gamma f}\,\right) \tag{35}$$

  Note that $u^{(k+1)}$ is nonnegative.

- Step 2: Here one has to solve an 'ordinary' L_2-TV minimization problem. This can be realized in different ways, e.g., by the simple gradient descent reprojection algorithm in [15] or by faster multistep algorithms [6, 50, 63].

Parallel to this paper, the splitting (II) was proposed *for the SO-model* by Bioucas-Dias and Figueiredo in [7]. For the function Ψ of the SO-model, Step 1 can also be decoupled and a componentwise analytical solution can be given in terms of the Lambert W function, see, e.g., [18]. However, the authors in [7] suggested better to apply a Newton method here. In connection with blurred Poissonian images this kind of splitting was used by the same authors in their so-called PIDAL (Poissonian image deconvolution by augmented Lagrangian) algorithm [30].

The ADMM for (III) reads

$$\begin{aligned} v^{(k+1)} &= \operatorname*{argmin}_{v \in \mathbb{R}^n} \{\frac{1}{2\gamma}\| \begin{pmatrix} b_1^{(k)} \\ b_2^{(k)} \end{pmatrix} + \begin{pmatrix} I \\ D \end{pmatrix} v - \begin{pmatrix} u^{(k)} \\ d^{(k)} \end{pmatrix} \|_2^2\}, \\ \begin{pmatrix} u^{(k+1)} \\ d^{(k+1)} \end{pmatrix} &= \operatorname*{argmin}_{u \in \mathbb{R}^n, d \in \mathbb{R}^{mn}} \{\Psi(u) + \lambda\phi(d) + \frac{1}{2\gamma}\| \begin{pmatrix} b_1^{(k)} \\ b_2^{(k)} \end{pmatrix} + \begin{pmatrix} I \\ D \end{pmatrix} v^{(k+1)} - \begin{pmatrix} u \\ d \end{pmatrix} \|_2^2\}, \\ b^{(k+1)} &= b^{(k)} + \begin{pmatrix} I \\ D \end{pmatrix} v^{(k+1)} - \begin{pmatrix} u^{(k+1)} \\ d^{(k+1)} \end{pmatrix} \end{aligned}$$

and requires in

- Step 1: the solution of a linear system of equations

$$v^{(k+1)} = (I + D^{\mathrm{T}}D)^{-1}\left(u^{(k)} - b_1^{(k)} + D^{\mathrm{T}}(d^{(k)} - b_2^{(k)})\right).$$

- Step 2: Here the minimization with respect to u and d can be decoupled. As already discussed above, $u^{(k+1)}$ can be computed componentwise by

$$u^{(k+1)} = \frac{1}{2}\left(b_1^{(k)} + v^{(k+1)} - \gamma + \sqrt{\left(b_1^{(k)} + v^{(k+1)} - \gamma\right)^2 + 4\gamma f}\right),$$

where nonnegativity is ensured and $d^{(k+1)}$ follows by the coupled shrinkage

$$d^{(k+1)} = T_{\gamma\lambda}(b_2^{(k)} + Dv^{(k+1)}).$$

This third splitting has the advantage that it requires no inner loops. It was already mentioned in [27] that 'multiple' splittings as (III) can be useful in various applications.

5 Numerical Results

For our numerical examples we use MATLAB implementations. All images are depicted in the interval $[0, 255]$. We restrict our attention to multiplicative noise η which follows a Gamma distribution with density function

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

Hence, η has mean 1 and standard deviation $\sigma = 1/\sqrt{L}$.
In our regularization terms we use the following matrices D:

i) For the discrete TV functional we use

$$D := \begin{pmatrix} I \otimes D_0 \\ D_0 \otimes I \end{pmatrix}, \quad D_0 := \begin{pmatrix} -1 & 1 & & & \\ 0 & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & & 0 \end{pmatrix}, \tag{36}$$

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: Initially, we start with a zero weight matrix w. For every image pixel $i = (i_1, i_2)$, we compute for all $j = (j_1, j_2)$ within a search window of size $\omega \times \omega$ around i the distances

$$d_a(i,j) := \sum_{t_1=-\lceil\frac{p-1}{2}\rceil}^{\lceil\frac{p-1}{2}\rceil} \sum_{t_2=-\lceil\frac{p-1}{2}\rceil}^{\lceil\frac{p-1}{2}\rceil} G_a(t_1,t_2)\big(h(i_1+t_1, i_2+t_2) - h(j_1+t_1, j_2+t_2)\big)^2, \tag{37}$$

where $h := \log(f)$ and G_a denotes the discretized, normalized Gaussian with standard deviation a. We refer to p as the patch size. Then, for given $\tilde{m}$ we select $k \leq \tilde{m}$ so-called 'neighbors' $j \neq i$ of i for which $d_a(i,j)$ takes the smallest values and set $w(i,j) = w(j,i) := 1$. By setting $w(i,j) = w(j,i)$ it happens that several weights $w(i,\cdot)$ are already non-zero before we reach pixel j. To avoid that the number of non-zero weights becomes too large, we choose only $k = \min\{\tilde{m}, 2\tilde{m} - l\}$ neighbors for l being the number of non-zero weights $w(i, \cdot)$ before the selection. Now, with regard to (4) and (17) we construct the matrix $D \in \mathbb{R}^{mn,n}$ with $m = 2\tilde{m}$ so that D 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.

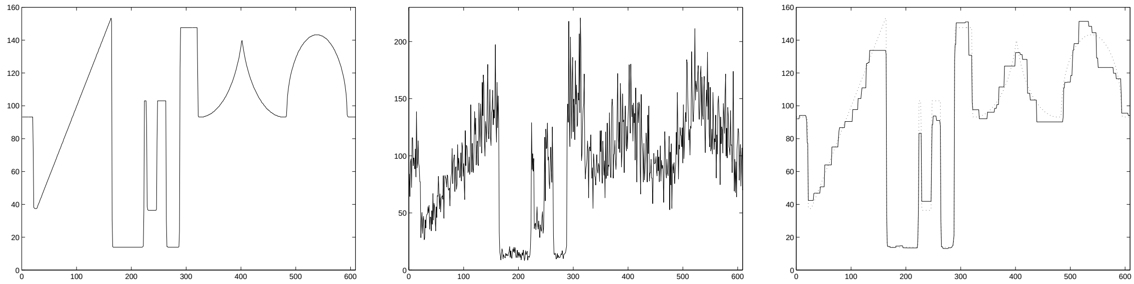

Figure 2: *Left*: Original signal. *Middle*: Noisy signal with Gamma noise of standard deviation 0.2. *Right*: Denoised signal by the I-divergence - TV model with $\lambda = 0.52$.

Our first example in Fig. 2 shows a restored 1D signal. The signal and noise level were chosen in accordance with the experiments in [59]. Although the I-divergence - TV model was originally designed for Poisson noise, it restores the signal quite nicely except for the usual staircasing artifacts typical for TV regularization.

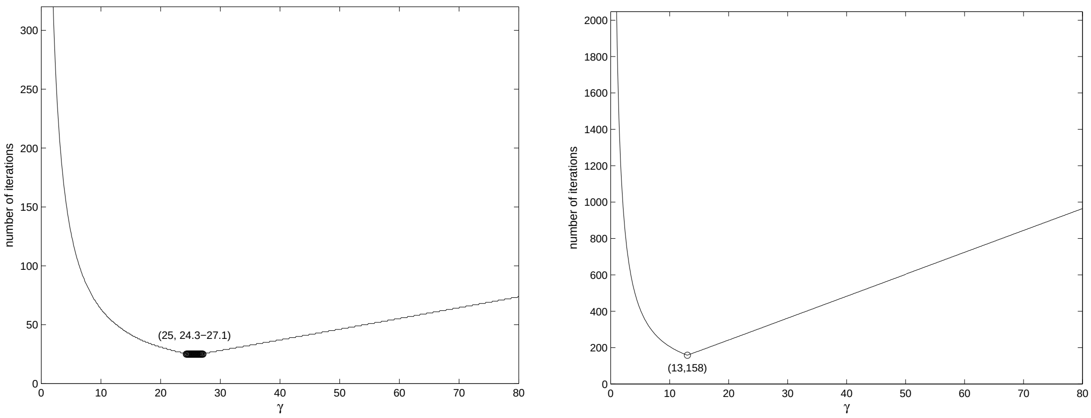

Figure 3: Influence of γ on the number of iterations needed by the ADMM corresponding to splitting (I) (with $\tau = 0$) to compute the restored signal u on the right of Fig. 2 up to a maximal error per pixel of $\varepsilon = 1$ (left) and $\varepsilon = 0.01$ (right).

To illustrate the *influence of the parameter* γ on the speed of the different algorithms we included Fig. 3. For the noisy signal given in Fig. 2 (left) we iterated the ADMM corresponding to splitting (I) with $\lambda = 0.52$ and $\tau = 0$ until $\max_i |\hat{u} - u_i^{(k)}| < \varepsilon$ for a sufficiently converged reference result $\hat{u}$. In all our experiments it was sufficient to iterate the inner loop only once per outer iteration. The plots show the necessary numbers of iterations for different values of γ. Interestingly, the curves increase linearly on the right of the minima. On the left the shape resembles $1/x$. Moreover, the optimal value of γ depends on the chosen ε: For a smaller value ε the optimal γ is also smaller. These observations could also be confirmed by 2D examples and are equally true for the other two splittings.

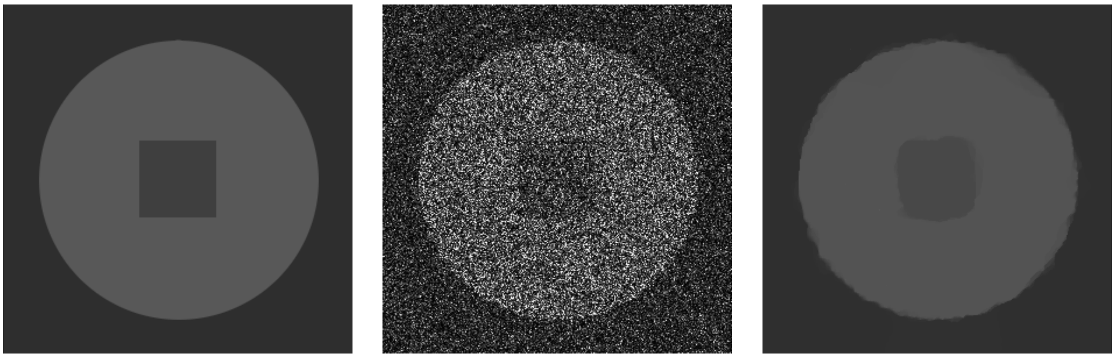

Figure 4: *Left*: Reproduction of a test image in [4]. *Middle*: Noisy image corrupted by Gamma noise for $\sigma = 1$ ($L = 1$). *Right*: Denoised image by the I-divergence - TV model with $\lambda = 2.5$.

Our next example in Fig. 4 illustrates the performance of the I-divergence - TV model in 2D. Here, the image on the left is a reproduction of a synthetic test image in [4]. Compared to the results given in [4] our denoising result looks very promising.
Next, we applied the I-divergence - TV model and the I-divergence - NL-means model to restore a part of the 'Barbara' image corrupted by Gamma noise of standard deviation 0.2 ($L = 25$). The figure and the noise level were chosen to keep the experiments comparable with the ones in [59]. As expected, the denoised image by the I-divergence - NL-means model is significantly better than the result obtained by the I-divergence - TV model due to the semi-local adaptivity of the NL-means matrix D.
With respect to the *computational speed* of the different algorithms, the ADMM for (I) with $\tau = 0$ and the ADMM for (III) outperform in general the algorithm corresponding to (II) for optimized γ. This is because in all our tests it took much more time to solve a complete L_2-TV problem in each iteration than to compute the solution of one of the involved linear systems of equations. The performance of the ADMMs corresponding to (I) and (III) depends mainly on the applied solver. In general, we used a CG method. Note that the involved matrices are always sparse. For the I-divergence - TV model with D defined in (36), the matrix $I + D^{\mathrm{T}}D$ occurring in the ADMM belonging to (III) has the advantage of being efficiently invertible by the discrete cosine transform. In this case, the ADMM for (III) performs best.
To present further examples, Fig. 6 shows the restoration results for a part of the 'Cameraman' image with a higher noise level of standard deviation 0.5 ($L = 4$) and in Fig. 7 we applied our methods also to a real-world SAR image. Note that by the construction of the NL-means matrix D the weights are always computed with respect to the logarithm of the corrupted image. While we were preparing the final version of this paper, we got to know about the parallel work [22] of Deledalle et al., who propose an iterated non-local means filter with probabilistic patch-based weights adapted to SAR images.

Figure 5: *Top*: Original image (left) and noisy version corrupted by Gamma noise of standard deviation 0.2 ($L = 25$) (right). *Bottom*: Denoised image by the I-divergence - TV model with $\lambda = 0.2$ (left) and the result by the I-divergence - NL-means model with $\lambda = 0.12$, $a = 3$, $p = 9$, $\omega = 11$ and $\tilde{m} = 5$ (right).

6 Conclusion

We have examined theoretically and numerically the suitability of the I-divergence - TV model as well as the I-divergence - NL-means model for restoring images contaminated by multiplicative Gamma noise. Furthermore, we showed how to efficiently solve the involved minimization problems by applying Douglas-Rachford splitting techniques, resp. the alternating direction method of multipliers. Numerical speed comparisons as well as theoretical examinations of the influence of the parameter γ on the ADMM, more sophisticated NL-means constructions and anisotropic functionals for handling non-additive noise are topics of future research.

Acknowledgment Many thanks to M. Grasmair (University of Vienna) for making the draft of [41] available, to M. Sigelle and F. Tupin (Telecom ParisTech, Paris) for providing us with the image in Fig. 7 and to S. Setzer for fruitful discussions.

Figure 6: *Top:* Original image (left) and noisy version corrupted by multiplicative Gamma noise of standard deviation 0.5 ($L = 4$) (right). *Bottom:* Restored image by the I-divergence - TV model with $\lambda = 0.9$ (left) and the result by the I-divergence - NL-means model with $\lambda = 0.6$, $a = 1.5$, $p = 5$, $\omega = 21$ and $\tilde{m} = 5$ (right).

References

[1] S. Aja-Fernández and C. Alberola-López. On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering. *IEEE Transactions on Image Processing*, 15(9):2694–2701, 2006.

[2] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhikers Guide*. Springer, 3rd edition, 2007.

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

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

[5] J.-P. Aubin. *Optima and Equilibria: An Introduction to Nonlinear Analysis*, volume 140 of *Graduate Texts in Mathematics*. Springer, Berlin - Heidelberg - New York, 2nd edition, 1998.

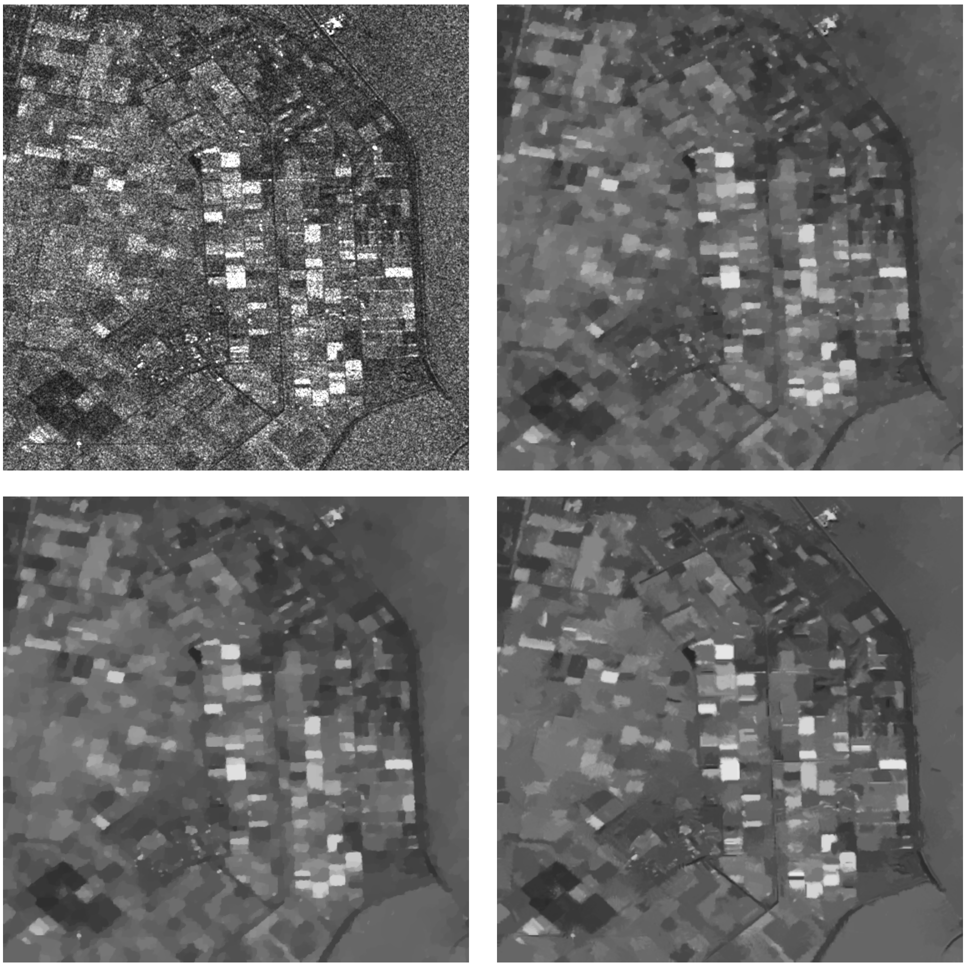

Figure 7: *Top left:* Original multi-look SAR image (copyright by [28]). *Top right and bottom left:* Restored images by the I-divergence - TV model with $\lambda = 0.4$ (top right) and $\lambda = 0.5$ (bottom left). *Bottom right:* Result by the I-divergence - NL-means model with $\lambda = 0.2$, $a = 3$, $p = 9$, $\omega = 13$ and $\tilde{m} = 5$.

[6] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 2009. To appear.

[7] J. M. Bioucas-Dias and M. A. T. Figueiredo. Total variation restoration of speckled images using a split-Bregman algorithm. IEEE International Conference on Image Processing, Cairo, Egypt, 2009.

[8] A. Braides. *Gamma-Convergence for Beginners*, volume 22 of *Oxford Lecture Series in Mathematics and Its Applications*. Oxford University Press, Oxford, 2002.

[9] E. Bratsolis and M. Sigelle. Fast SAR image restoration, segmentation and detection of high-reflectance regions. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2890–2899, 2003.

[10] 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.

[11] C. Brune, A. Sawatzky, F. Wübbeling, T. Kösters, and M. Burger. An analytical view of EM-TV based methods for inverse problems with poisson noise. Preprint, University of Münster, 2009.

[12] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 60–65, 2005.

[13] Y. Censor and A. Lent. An interval row action method for interval convex programming. *Journal of Optomization Theory and Applications*, 34:321–353, 1981.

[14] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

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

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

[17] P. L. 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.

[18] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, (5):329–359, 1996.

[19] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot. Nonlocal means-based speckle filtering for ultrasound images. *IEEE Transactions on Image Processing*, 18(10):2221–2229, October 2009.

[20] I. Csiszár. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2032–2066, 1991.

[21] G. Dal Maso. *An Introduction to Γ-Convergence*. Birkhäuser, Boston, 1993.

[22] C.-A. Deledalle, L. Denis, and F. Tupin. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Transactions on Image Processing*, 18(12):2661–2672, 2009.

[23] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82(2):421–439, 1956.

[24] S. Durand, J. Fadili, and M. Nikolova. Multiplicative noise cleaning via a variational method involving curvelet coefficients. In X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods in Computer Vision*, volume 5567 of *LNCS*, pages 282–294. Springer, 2009.

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

[26] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. SIAM, 1999.

[27] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. CAM Report 09-31, UCLA, 2009.

[28] European Space Agency (ESA). Multilook SAR image.

[29] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Propoerties of Functions*. Studies in Advanced Mathematics. CRC Press, 1992.

[30] M. A. T. Figueiredo and J. M. Bioucas-Dias. Deconvolution of Poissonian images using variable splitting and augmented Lagrangian optimization. In *IEEE Workshop on Statistical Signal Processing*, Cardiff, 2009.

[31] I. Fonseca and G. Leoni. *Modern Methods in the Calculus of Variations: L^p Spaces*. Springer Monographs in Mathematics. Springer, 2007.

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

[33] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary–Value Problems*, volume 15 of *Studies in Mathematics and its Applications*, chapter 9, pages 299–331. North–Holland, Amsterdam, 1983.

[34] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.

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

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

[37] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2008.

[38] R. Glowinski and A. Marrocco. Sur l'approximation par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 9(2):41–76, 1975.

[39] T. Goldstein and S. Osher. The Split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[40] 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.

[41] M. Grasmair. A coarea formula for anisotropic total variation regularisation. Preprint to appear, University of Vienna, 2009.

[42] Y.-M. Huang, M. K. Ng, and Y.-W. Wen. A new total variation method for multiplicative noise removal. *SIAM Journal on Imaging Sciences*, 2(1):20–40, 2009.

[43] C. Kervrann, J. Boulanger, and P. Coupé. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *LNCS*, pages 520–532. Springer, 2007.

[44] 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, 2007.

[45] A. Kryvanos, J. Hesser, and G. Steidl. Nonlinear image restoration methods for marker extraction in 3D fluorescent microscopy. In C. A. Bouman and E. L. Miller, editors, *Computational Imaging III, Proc. SPIE*, volume 5674, pages 432–443, 2005.

[46] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):165–177, 1985.

[47] J.-S. Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(2):165–168, 1980.

[48] 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.

[49] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. Monographs on Mathematical Modeling and Computation. SIAM, Philadelphia, 2001.

[50] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

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

[52] E. Resmerita, H. W. Engl, and A. N. Iusem. The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis. *Inverse Problems*, 23(6):2575–2588, 2007.

[53] R. T. Rockafellar. Integral functions, normal integrands and measurable selections. In J. P. Gossez, E. J. Lami Dozo, J. Mawhin, and L. Waelbroeck, editors, *Nonlinear Operators and the Calculus of Variations*, volume 543 of *Lecture Notes in Mathematics*, pages 157 – 207. Springer, 1976.

[54] 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 Set Methods in Imaging, Vision, and Graphics*, pages 103–119. Springer, 2003.

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

[56] A. Sawatzky, C. Brune, F. Wübbeling, T. Kösters, K. Schäfers, and M. Burger. Accurate EM-TV algorithm in PET with low SNR. In *IEEE Nuclear Science Symposium Conference Record*, pages 5133–5137, 2008.

[57] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, 2009.

[58] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods in Computer Vision*, volume 5567 of *LNCS*, pages 464–476. Springer, 2009.

[59] 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.

[60] G. Steidl. A note on the dual treatment of higher-order regularization functionals. *Computing*, 76(1-2):135–148, 2006.

[61] G. Steidl and T. Teuber. Anisotropic smoothing using double orientations. In X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods in Computer Vision*, volume 5567 of *LNCS*, pages 477–489. Springer, 2009.

[62] 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.

[63] P. Weiss, L. Blanc-Féraud, and G. Aubert. Efficient schemes for total variation minimization under constraints in image processing. *SIAM Journal on Scientific Computing*, 31(3):2047–2080, 2009.

[64] M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: a link between diffusion filtering and wavelet shrinkage. *Applied and Computational Harmonic Analysis*, 24(2):195–224, 2008.

[65] 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.

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

[67] D. Zhou and B. Schölkopf. Regularization on discrete spaces. In W. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition, Proceedings of the 27th DAGM Symposium*, volume 3663 of *Lecture Notes in Computer Science*, pages 361–368. Springer, 2005.