Smoothing the finite differences defining the Non-local Total Variation and application in image restoration

Zhi Li, François Malgouyres, Tieyong Zeng

Noname manuscript No.
(will be inserted by the editor)

Smoothing the finite differences defining the Non-local Total Variation and application in image restoration

Zhi Li · François Malgouyres · Tieyong Zeng

Received: date / Accepted: date

Abstract In the usual non-local variational models, such as the non-local total variations (NLTV), the image is regularized by minimizing an energy term that penalizes gray-levels discrepancy between some specified pairs of pixels. The pairs of pixels are interpreted as weighted edges joining nodes in a graph. The nodes correspond to the pixels of the image. The weights are ponderations. The graph structure is usually determined a-priori or, sometimes, is adjusted (most of the time empirically) during the iterative process.

In this paper, we study the possibility to impose some regularity to the graph. More precisely, we study a model in which two edges whose starting nodes are near and such that the starting and ending points form a parallelogram to carry similar weights. In order to do so, we minimize a function involving a regularization term, analogous to an H^1 term, on the graph edges. Doing so, the finite differences defining the image regularity depend on their environment. They are therefore more stable. The model is also better suited to solve inverse problems for which the design of the graph is not straightforward (one can think of computerized tomography problems, reconstruction from random measures, inpainting...)

We provide all the details necessary for the implementation of a PALM algorithm with proved convergence. We illustrate the ability of the model to restore relevant unknown edges from the neighboring edges on an image inpainting problem. We also argue on inpainting and denoising problems that the model better recovers thin structures.

Keywords non-local regularization · proximal alternating linearized minimization · nonconvex minimization · total variation · image restoration

Mathematics Subject Classification (2000) 49N45 · 65K10 · 90C26

Z. Li
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China,
E-mail: zhi_li@life.hkbu.edu.hk

F. Malgouyres
University of Toulouse, CNRS, Toulouse, France,
E-mail: Francois.Malgouyres@math.univ-toulouse.fr.

T. Zeng
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China,
E-mail: zeng@hkbu.edu.hk. TZ is partially supported by NSFC 11271049, RGC 12302714 and RFGs of HKBU.

1 Introduction

1.1 The Image Restoration Problem

Given an image domain $\mathcal{P}$ ($\mathcal{P}$ can be a sub-domain of $\mathbb{R}^2$ or a lattice of pixels), we consider a degraded image $I \in \mathbb{R}^{\mathcal{P}}$. We assume that I is obtained from a perfect unknown image $\overline{u} \in \mathbb{R}^{\mathcal{P}}$ by

$$I = H\overline{u} + b,$$

where H is a known linear operator and $b \in \mathbb{R}^{\mathcal{P}}$ is an error term. In the typical situations we have in mind b is an iid noise with standard deviation $\sigma > 0$.

The purpose of the restoration model studied in this paper is to retrieve an image close to $\overline{u}$, using only I and H. For instance, in image inpainting, the operator H maps any image $u \in \mathbb{R}^{\mathcal{P}}$ onto an

image $Hu \in \mathbb{R}^{\mathcal{P}}$ satisfying

$$(Hu)_p = \begin{cases} 0 & \text{, if } p \in \mathcal{P}', \\ u_p & \text{, otherwise,} \end{cases} \tag{1}$$

where $\mathcal{P}'$ is the domain to inpaint.

For deblurring, the operator H is typically the convolution operator. For the purpose of image zooming the operator H is the composition of a convolution and a subsampling operator (the latter can also be written under the form (1)) and so on and so forth. Note that, to simplify notations, we use the same domain $\mathcal{P}$ for the perfect image $\overline{u}$ and the degraded image I. This need not be the case in general.

Finally, for simplicity, we assume that $\mathcal{P}$ is periodized, so that for any $(p, q) \in \mathcal{P}^2$ the sum $p + q$ is defined in such a manner that $p+q \in \mathcal{P}$. We also assume that $\mathcal{P}$ is such that $p + \mathcal{P} = \mathcal{P}$, whatever $p \in \mathcal{P}$.

1.2 Review of Image Restoration Models

Total variation (TV) has drawn much attention since Rudin, Osher and Fatemi (ROF) proposed their celebrated model [39] for denoising (i.e. $H = Id$). The variational formulation can be written as

$$\underset{u \in \Omega}{\operatorname{argmin}} \quad |\nabla u| + \lambda \|u - I\|^2, \tag{2}$$

where Ω is either $\mathbb{R}^{\mathcal{P}}$ (in the discrete setting) or a space of the functions of bounded variation. When used as a regularization term, the total variation $|\nabla u|$ has the advantage to preserve sharp edges and the drawbacks to create homogeneous regions (staircasing) and to over-smooth the textured regions. This is due to the fact that textures are treated as singularities by the TV term which only sees local information [27]. This regularization has later been used to solve many linear inverse problems such as inpainting [4, 41].

In another direction, generalizing the Yaroslavsky neighborhood filter [46], the non-local means (NLM) of Buades, Coll and Morel [8] takes the following form (we write it in the continuous domain):

$$NLM_I(p) = \frac{1}{C(p)} \int_{\mathcal{B}} v_I(p, p') I(p') dp',$$

where $\mathcal{B}$ is a search window, the weights are provided by

$$v_I(p, p') = \exp\left(-\frac{D(p, p')}{h^2}\right), \tag{3}$$

where, for a patch $\hat{\mathcal{B}}$ (usually a square centered around the origin),

$$D(p, p') = \int_{\hat{\mathcal{B}}} G_a(q) \left|I(p + q) - I(p' + q)\right|^2 dq$$

and the normalizing factor

$$C(p) = \int_{\mathcal{B}} v_I(p, p') dp'.$$

The Gaussian kernel G_a with standard deviation a ponderates pixels in the 'patch' area and h is a filtering parameter. The idea behind NLM is to restore the pixel in the contaminated image I by a weighted average of the nearby similar (in the patch sense) pixels. A key point is that, for every pixel, the weights of its neighbors is calculated based on 'patch' resemblance instead of intensity values similarity. Another key point is that the pixels p' used in the average do not have to be close to p (thus the name non-local). Notice that a similar idea, coming with a precise statistical analysis, has been proposed under the name "Discrete Universal Denoiser" [45].

Considering a thin structure (typically a smooth curve) in a uniform background, we see in the formula that the weights v_I does not differentiate well the pixels in the thin structure from pixel in the background. The thin structure indeed leaves in a domain whose measure is small (at the limit, the measure of the domain is 0). It provides a negligible contribution to $D(p, p')$. As a consequence, NLM tends to underestimate the weights of the pixels in the thin structure and, as a consequence, it tends to erase the thin structure. This is especially true for large patch size and may lead to a substantial underestimation of the similarity between two pixels in the thin structure [20].

Moreover, exemplar-based approaches [30, 43] also use patches to produce striking inpainting results. These algorithms have been later interpreted using a probabilistic point of view. In particular, Levina and Bickel [30] provided a theoretical background to explain their ability to synthesize texture. Aujol, Ladjal and Masnou [3] expressed the exemplar-based methods in a non-local variational framework and studied their ability to retrieve geometric features. However, it is difficult to adapt NLM and other patch based approaches for difficult image restoration tasks involving highly structured noise or (even worse) when the operator H is not invertible and makes the recovery of similar pixels difficult.

Inspired by the advantages of NLM, Gilboa and Osher [24, 25] introduced a non-local variational

model to better recover textures. In [25], they considered the non-local total variation (NLTV) model defined (in the continuous domain) by

$$J_{NLTV}(u) + \lambda \|u - I\|^2,$$

where the NLTV term $J_{NLTV}(u)$ is defined by

$$\int_{\mathcal{P}} \sqrt{\int_{\mathcal{B}} (u(p) - u(p+q))^2 \, v(p,q) dq dp}. \tag{4}$$

In [25], in practical applications, they calculated the weights v according to patch distance and only used the 10 most similar pixels in (the discretized) $\mathcal{B}$. They also mentioned binary weights. They adapted the celebrated Chambolle's projection algorithm [11] to optimize a discretization of the NLTV model. A more general model, involving p-Laplace operator on graphs, have been proposed in [21]. It also makes connections between the variational approaches and other non-local approaches.

Considering applications such as compressive sensing, when the low sampling rate may lead to a bad initial guess of the weight function, some authors [38, 47] suggested to update the weights during the iterations to improve them. For example, Zhang, et al. [47] employed patches and updated the weights by recomputing the initial step (3) with the current estimate of the clean image. A similar update is used in [1,38]. To trade off between the accuracy and computational costs, the authors only reevaluate the weights every few steps [47] and in practical implementation, they only use the first few largest weights in a local search window and set the rest of them to zero. Of course, because of these last "tricks", the algorithms may fail to find the optimal solution of the variational model which is not fully satisfactory. They however provide interesting restoration results.

The model studied in this paper is inspired by the earlier work of Peyre, et al. [37, 38]. In [37], they considered the weights v defined in the NLM and updated v using (3) with the current estimate of the image. The authors mentioned the difficulty for setting the filtering parameter h [37] since noise level evolves along the iterative process. In [38], the authors extended [37] and [47] by regularizing the graph v. The resulting model has two regularization terms, the first term $J_v(u)$ regularizes the image, while the second term $E(v)$ is the entropy of function v defining the underlying graph. Also, to better recover the textures of the images, the non-local regularization term was chosen to be the patch variation of $J_{NLTV}(u)$ in (4). In this variation, the difference of intensities $u(p) - u(p+q)$, in (4), is replaced by a patch distance. It has also been adopted in [1, 22, 36].

To conclude, non-local variational models [19, 25, 32, 38] have been widely used to restore images, because of their capability to utilize the information from many pixels having a similar neighborhood. However, when treated by these models, thin structures can be overlooked and are sometimes undetectable in a patch of the damaged image. These method might therefore erase these thin structures.

Also, the construction of the weights between the pixels is usually done according to some pixel similarity criterion in the degraded images I. When some weights are unknown (for instance when inpainting a missing domain or zooming an image), they are updated during the iterative process, according to an ad-hoc procedure [38]. More generally, these methods are difficult to apply and lack precise modeling when the output of the degradation operator H does not look like an image (think about the restoration an image from random measurements or Computerized Tomography.)

When compared to these methods, the model studied in this paper is a regularized non-local total variation (RNLTV) model. It shows the advantages to both: - better restore thin structures; - allow the construction of non-local models for difficult inverse problems involving degradation operator whose output does not resemble the input image.

1.3 Overview of the Paper

In Section 2, we present the RNLTV model. Section 3 details the proximal alternating linearized minimization (PALM) algorithm [6] minimizing the proposed model. We also provide in this section the gradient of the different terms, show that the gradients are Lipschitz continuous and provide their Lipschitz constants. The numerical experiments in Section 4 demonstrate the efficiency of the proposed method on denoising, inpainting and zooming experiments.

2 The Studied Model

2.1 Framework and Model Description

From now on, we discretize the problem and consider a lattice $\mathcal{P}$. We consider a set $\mathcal{B} \subset \mathcal{P}$. In the typical situation we have in mind, $\mathcal{B}$ is a square

centered at the origin with radius[1] r. Throughout the paper, we call patch or window a set $p+\mathcal{B}$ for some $p \in \mathcal{P}$.

We also denote the cardinality of any finite set C by $|C|$. Moreover, for any finite dimensional vector space V (in practice we have either $V = \mathbb{R}^{\mathcal{P}}$, $V = \mathbb{R}^{\mathcal{P}\times\mathcal{B}}$ or $V = \mathbb{R}^{\mathcal{B}}$) and any positive definite operator L mapping V into itself, we consider the weighted norm [17]

$$\|x\|_L = \langle x, Lx\rangle^{\frac{1}{2}}, \qquad \forall x \in V,$$

where $\langle ., . \rangle$ denotes the usual scalar product in V. When L is an identity matrix ($L = Id$), we simply denote $\|x\|$ instead of $\|x\|_{Id}$.

Given these notations, we define, for any $p \in \mathcal{P}$, weights $v^p = (v^p_q)_{q\in\mathcal{B}} \in \mathbb{R}^{\mathcal{B}}$, where each weight v^p_q is the cost for the dissimilarity $|u_p - u_{p+q}|$. Said differently, $v^p \in \mathbb{R}^{\mathcal{B}}$ can be seen as the parameters of a finite difference operator defined at the pixel p. In practice we will force these weights $v^p \in \mathbb{R}^{\mathcal{B}}$ to belong to the set

$$\mathcal{U} = \left\{ v \in \mathbb{R}^{\mathcal{B}}_+, \sum_{q\in\mathcal{B}} v_q = 1 \right\},$$

where $\mathbb{R}_+$ are the non-negative reals. This constraints guarantees that every pixel of the image contributes similarly to the overall cost. We denote the collection of all the weights in the image by $v = (v^p)_{p\in\mathcal{P}} \in \mathcal{U}^{\mathcal{P}}$.

Given v, we define the non-local total variation of $u \in \mathbb{R}^{\mathcal{P}}$ by

$$TV\,(v,u) = \sum_{p\in\mathcal{P}} \Psi_\mu \left(\sqrt{\sum_{q\in\mathcal{B}} v^p_q (u_p - u_{p+q})^2} \right), \tag{5}$$

where $\mu > 0$ and Ψ_μ is the Moreau envelope of the absolute value function, also called Huber function [5, page 368]:

$$\psi_\mu(t) = \begin{cases} |t| - \frac{\mu}{2} & \text{, if } |t| \geq \mu, \\ \frac{t^2}{2\mu} & \text{, otherwise.} \end{cases}$$

Note that (4) is a continuous analogue of (5) without the Huber function Ψ_μ. Moreover, we consider, for any $v \in \mathcal{U}^{\mathcal{P}}$, the regularity criterion R defined by

$$R(v) = \gamma \sum_{p\in\mathcal{P}} \sum_{p'\in\mathcal{N}} \|v^p - v^{p+p'}\|^2, \tag{6}$$

where γ is a non-negative parameter and $\mathcal{N}$ is a small neighborhood[2] of 0.

Finally, we consider a data fidelity term D. This term either takes the form

$$D(u) = \lambda \|Hu - I\|^2,$$

for some parameter $\lambda \geq 0$, or, for a parameter $\tau \geq 0$,

$$D(u) = \chi_{|\|Hu-I\|^2 \leq \tau}(u),$$

where the indicator function $\chi_{|C}$ is 0 if the condition C is satisfied and infinity otherwise. The advantage of the latter formulation is that τ can be deduced from the noise standard deviation σ (A standard choice is indeed $\tau = |\mathcal{P}|\sigma^2$).

The restored image we are interested in is the argument u of a minimizer of

$$\underset{u\in\mathbb{R}^{\mathcal{P}}, v\in\mathcal{U}^{\mathcal{P}}}{\operatorname{argmin}} \quad R(v) + TV\,(v,u) + D(u). \tag{7}$$

Although the parameters are not presented explicitly in (7), the model depends on the parameters μ, γ and either λ or τ.

In the problem (7), the constraint on v is compact and, unless in very particular counter-example, we expect the objective function to be coercive in u. The objective function being also lower semi-continuous, the problem (7) has a solution. We will see however that TV and R are differentiable and have Lipschitz gradient (see Proposition 1, 2 and 3). The main difficulty concerning the numerical resolution of (7) is that, because of the term $TV(u,v)$, the objective function is not convex.

2.2 Comments on the Model

The connectivity principle [28, 33] states that human prefers to see that disjoint edges are connected through the missing domains. It is well-known that the classic inpainting models including TV models [4,41], have difficulty in connecting the 'slim' objects [13,14], thus inpainting methods involving higher order information [13,40], such as curvature, were proposed to improve the results. However, if the size of the unknown regions separating slim objects get too large, then even the fourth-order PDE-based model (TV-H^{-1}) [40] will leave the thin structures broken. From Figure 1, we can see that TV model do not connect any lines, TV-H^{-1} can connect the middle one. However, it is

[1] We use $r = 5$ in the experiments.

[2] In the experiments, we consider 4-connectivity: $\mathcal{N} = \{(1,0),(0,1)\}$.

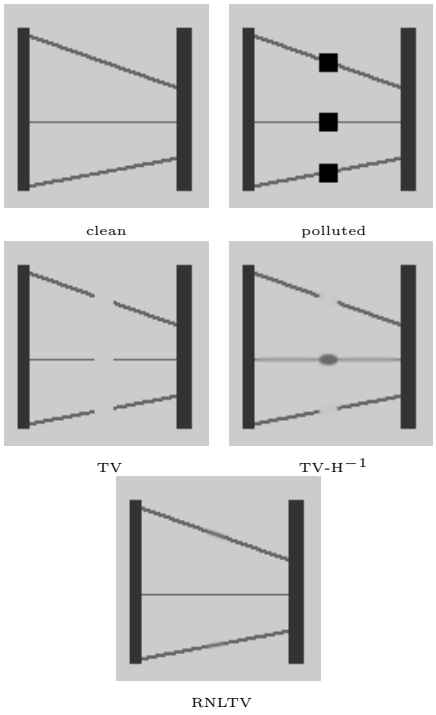

Fig. 1 A synthetic image with thin lines contaminated by 11×11 missing squares. In the case of large missing areas both the TV [41] and the fourth-order PDE-based inpainting model (TV-H^{-1}) [40] fails to follow the Connectivity Principle properly, but our RNLTV can connect those disjoint thin lines.

an ellipse with wrong color that fills in the interval between the disjoint lines. The zero curvature on the line could be the reason leading to the bizarre shape. Our RNLTV connects the broken lines and exactly restores the middle line.

We focus on maximizing the ability of the NLTV model to restore thin structures by allowing the weights v in the non-local TV to change during the iterations. Note that the conventional NLTV is essentially a local method [26], due to the limited size of the window $\mathcal{B}$ and the patch, NLTV can not use the global information in the whole image to further restore the polluted image. Similarly, the NLM [7] was proposed as a global method to search the whole image for similar pixels [29], but is implemented in a 'semi-local' way [9], that means the searchings are restricted to a local window. Additionally, increasing the size of the local search region will also cause heavy computational burden, and can not always lead to improvements [24, 29].

Moreover, despite the parameter γ, the update of the weights v does not rely on difficult and empirical parameter tuning. For example, in the denoising case, the noise level decreases during the iterative process and since, when updated with (3), the filtering parameter h depends on the noise level, one should normally adapt h according to the iteration number to get more accurate weights [37, 38].

3 Implementing the PALM Algorithm

In the last couple of years, alternating minimization algorithms [2, 6, 15, 16, 17, 35] were developed to solve optimization problems over multiple variables. A proof of convergence in the non-convex setting for an alternating variant of the forward-backward algorithm, tagged PALM, has been proposed in [6]. It is perfectly adapted to the structure of (7). A preconditioned version has also been studied in [17].

Below, we provide the details for implementing the PALM algorithm generating a sequence that converges to a stationary point of (7). First, we give a few notations and describe the algorithm in Section 3.1. Next, we derive the components of the algorithm. Section 3.2, 3.3 and 3.4 are devoted to the calculation of the gradients of the regularization terms in (7) and there Lipschitz constants. Then we discuss the proximal operator $\text{prox}_L^D(u')$ in Section 3.5 in detail.

3.1 Notations and Overview of the Algorithm

We denote $\nabla_u TV(v,u)$ and $\nabla_v TV(v,u)$ as the partial gradient of $TV(.,.)$ with regards to the variable u and v respectively. The PALM algorithm adapted to (7) is described in Algorithm 1. In summary, the Algorithm 1 alternates forward-backward iterations for each one of the two block of variables: u and v.

For any $v \in \mathcal{U}^{\mathcal{P}}$ and any $u \in \mathbb{R}^{\mathcal{P}}$, we assume that we know a linear operator L and constants l and l' (for the detailed calculations of these quantities see Section 3.2, 3.3 and 3.4) satisfying the majorization conditions [17] described below:

- For any $v \in \mathcal{U}^{\mathcal{P}}$, L is such that there exists $m > 0$ satisfying

  $$\|u\|_L \geq m\|u\|, \qquad \forall u \in \mathbb{R}^{\mathcal{P}},$$

  and for any u and $u' \in \mathbb{R}^{\mathcal{P}}$

$$\begin{aligned} TV(v,u') \leq\ & TV(v,u) \\ & + \langle \nabla_u TV(v,u), u'-u \rangle \\ & + \frac{1}{2}\langle u'-u, L(u'-u) \rangle . \end{aligned} \tag{8}$$

– For any $u \in \mathbb{R}^{\mathcal{P}}$, $l \geq 0$ and $l' \geq 0$ are such that $l + l' > 0$ and for any v and $v' \in \mathcal{U}^{\mathcal{P}}$

$$\begin{aligned} TV\left(v', u\right) \leq & TV\left(v, u\right) \\ & + \left\langle \nabla_v TV\left(v, u\right), v' - v \right\rangle \\ & + \frac{l}{2}\|v' - v\|^2, \end{aligned} \tag{9}$$

and

$$\begin{aligned} R(v') \leq & R(v) + \langle \nabla R(v), v' - v \rangle \\ & + \frac{l'}{2}\|v' - v\|^2. \end{aligned} \tag{10}$$

Moreover, the proximal operators used in Algorithm 1 are defined, for any $u' \in \mathbb{R}^{\mathcal{P}}$, by

$$\text{prox}_L^D\left(u'\right) = \underset{u \in \mathbb{R}^{\mathcal{P}}}{\text{argmin}}\ D(u) + \frac{1}{2}\|u - u'\|_L^2, \tag{11}$$

and for any $v' \in \mathcal{U}^{\mathcal{P}}$,

$$\text{prox}^{\chi_{|\mathcal{U}^{\mathcal{P}}}}\left(v'\right) = \underset{v \in \mathcal{U}^{\mathcal{P}}}{\text{argmin}} \|v - v'\|^2, \tag{12}$$

simply is the orthogonal projection onto $\mathcal{U}^{\mathcal{P}}$.

Algorithm 1: Overview of the algorithm

Input:
I: Degraded image
H: Linear operator to invert
$\mathcal{B}$: Support of the weights
$\mathcal{N}$: Support for the finite differences defining R
γ, μ, λ/τ: parameters
Output:
u: Restored image
v: Weights

begin
  Initialize the image u and v;
  while *not converged* **do**
    step 1: Update u:
    $u = \text{prox}_L^D\left(u - L^{-1}\nabla_u TV\left(v, u\right)\right)$
    step 2: Update v:
    $v = \text{prox}^{\chi_{|\mathcal{U}^{\mathcal{P}}}}\left(v - \frac{\nabla_v TV(v,u) + \nabla R(v)}{l + l'}\right)$

The Algorithm 1 has many advantages. In the context of our problem, it is shown in [17] that the sequence generated by (even an inexact version of) Algorithm 1 has good convergence properties:

– The objective function decays along the iterative process.
– The iterates converge to a stationary point of our problem.
– The sum over the iterative process of the distance between two successive iterates is finite.

All these properties are satisfied by the sequence generated by Algorithm 1. Finally, in order to implement the algorithm, we need to

– compute $\nabla_u TV\left(v, u\right)$ and construct the operator L (see Section 3.2);
– compute $\nabla_v TV\left(v, u\right)$ and the constant l satisfying (8) (see Section 3.3);
– compute $\nabla R(v)$ and the constant l' satisfying (9) (see Section 3.4);
– compute the proximal operator (11) (see Section 3.5);
– compute the proximal operator (12): Notice that this reduces to multiple projections onto simplex. The latter projection is well studied problem for which fast algorithms exists [18].

3.2 Calculation of $\nabla_u TV\left(v, u\right)$ and L

For any $v \in \mathcal{U}^{\mathcal{P}}$, we consider the following nonlocal analogue of the finite difference operator:

$$\begin{aligned} \mathbf{D}_v : \quad & \mathbb{R}^{\mathcal{P}} \longrightarrow \mathbb{R}^{\mathcal{P}\times\mathcal{B}} \\ & u = (u_p)_{p\in\mathcal{P}} \longmapsto \left(\sqrt{v_q^p}\ (u_p - u_{p+q})\right)_{(p,q)\in\mathcal{P}\times\mathcal{B}} \end{aligned}$$

We also denote, for $p \in \mathcal{P}$,

$$(\mathbf{D}_v u)_p = ((\mathbf{D}_v u)_{p,q})_{q\in\mathcal{B}} \in \mathbb{R}^{\mathcal{B}}$$

and $\|(\mathbf{D}_v u)_p\| = \sqrt{\sum_{q\in\mathcal{B}} (\mathbf{D}_v u)_{p,q}^2}$.

We obtain after some simple calculations that its adjoint is given, for any $w = (w_{p,q})_{(p,q)\in\mathcal{P}\times\mathcal{B}}$, by (see Appendix 6.1)

$$(\mathbf{D}_v^* w)_p = \sum_{q\in\mathcal{B}} \left(\sqrt{v_q^p}\ w_{p,q} - \sqrt{v_q^{p-q}}\ w_{p-q,q}\right),$$

for all $p \in \mathcal{P}$.

Moreover, if we consider, for $w \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}}$, the norm

$$\|w\|_{\infty,2} = \max_{p\in\mathcal{P}} \|w_p\|, \text{ where } w_p = (w_{p,q})_{q\in\mathcal{B}} \in \mathbb{R}^{\mathcal{B}}.$$

We can prove (see appendix 6.2) that for any $u \in \mathbb{R}^{\mathcal{P}}$ and any (fixed) $v \in \mathcal{U}^{\mathcal{P}}$

$$TV\left(v, u\right) = \max_{\|w\|_{\infty,2}\leq 1} \langle \mathbf{D}_v u, w\rangle - \frac{\mu}{2}\|w\|^2. \tag{13}$$

From this expression and using the Theorem 1 in [34] (reminded in Appendix 6.3), we can deduce the following proposition whose proof is detailed in Appendix 6.3. Notice that a similar statement and construction is given for the usual total variation in [44].

Proposition 1 *For any $u \in \mathbb{R}^{\mathcal{P}}$, we have*

$$\nabla_u TV\,(v,u) = \mathbf{D}_v^* w^*(u),$$

where $w^*(u) \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}}$ is the maximizer of (13) and is provided in closed form by:

$$w^*(u)_{p,q} = \begin{cases} \frac{(\mathbf{D}_v u)_{p,q}}{\mu} & \text{, if } \|(\mathbf{D}_v u)_p\| \leq \mu, \\ \frac{(\mathbf{D}_v u)_{p,q}}{\|(\mathbf{D}_v u)_p\|} & \text{, otherwise.} \end{cases}$$

Moreover, $u \to \nabla_u TV\,(v,u)$ is Lipschitz continuous with Lipschitz constant

$$l" = \frac{\sqrt{2}\sqrt{|\mathcal{B}|+1}}{\mu}.$$

As a consequence, we have for any u and $u' \in \mathbb{R}^{\mathcal{P}}$

$$\begin{aligned} TV\,(v,u') \leq\ & TV\,(v,u) \\ & + \langle \nabla_u TV\,(v,u)\,, u'-u \rangle \\ & + \frac{l"}{2}\|u'-u\|^2. \end{aligned} \tag{14}$$

A straightforward consequence of (14) is that for any diagonalizable linear operator G whose eigenvalues are all larger than 1 (and in particular for $G = Id$), the property (8) holds for $L = l"G$.

3.3 Calculation of $\nabla_v TV\,(v,u)$ and l

Let us first define, for any fixed $u \in \mathbb{R}^{\mathcal{P}}$, the linear operator $\mathbf{A}_u$:

$$\begin{aligned} \mathbb{R}^{\mathcal{P}\times\mathcal{B}} &\longrightarrow \mathbb{R}^{\mathcal{P}}, \\ (v_{p,q})_{(p,q)\in\mathcal{P}\times\mathcal{B}} &\longmapsto \left(\textstyle\sum_{q\in\mathcal{B}} v_{p,q}(u_p - u_{p+q})^2\right)_{p\in\mathcal{P}}. \end{aligned}$$

The interest for this operator comes from the fact that for any $v \in \mathcal{U}^{\mathcal{P}}$

$$TV\,(v,u) = \sum_{p\in\mathcal{P}} \Psi_\mu\left(\sqrt{(\mathbf{A}_u v)_p}\right).$$

Notice that, when $v \in \mathcal{U}^{\mathcal{P}}$, we have for any $p \in \mathcal{P}$

$$(\mathbf{A}_u v)_p \geq 0.$$

We easily obtain (see Appendix 6.4) that the adjoint $\mathbf{A}_u^*$ of $\mathbf{A}_u$ is given for any $w \in \mathbb{R}^{\mathcal{P}}$ by

$$(\mathbf{A}_u^* w)_{p,q} = w_p(u_p - u_{p+q})^2, \qquad \forall (p,q) \in \mathcal{P}\times\mathcal{B}.$$

Using these notations, we can prove (see Appendix 6.5) the following proposition.

Proposition 2 *For any fixed $u \in \mathbb{R}^{\mathcal{P}}$, we have for any $v \in \mathcal{U}^{\mathcal{P}}$*

$$\nabla_v TV\,(v,u) = \mathbf{A}_u^* u^*(v),$$

where for all $p \in \mathcal{P}$

$$(u^*(v))_p = \begin{cases} \frac{1}{2\sqrt{(\mathbf{A}_u v)_p}} & \text{, if } \sqrt{(\mathbf{A}_u v)_p} \geq \mu, \\ \frac{1}{2\mu} & \text{, if } \mu > \sqrt{(\mathbf{A}_u v)_p} \geq 0. \end{cases}$$

Moreover, the function $v \longmapsto TV\,(v,u)$ is concave over $\mathbb{R}_+^{\mathcal{P}\times\mathcal{B}}$ and (9) holds for $l = 0$.

3.4 Calculation of $\nabla R(v)$ and l'

Let us first define the linear operator B:

$$\begin{aligned} \mathbb{R}^{\mathcal{P}\times\mathcal{B}} &\longrightarrow \mathbb{R}^{\mathcal{P}\times\mathcal{B}\times\mathcal{N}}, \\ (v_{p,q})_{(p,q)\in\mathcal{P}\times\mathcal{B}} &\longmapsto (v_{p,q} - v_{p+p',q})_{(p,q,p')\in\mathcal{P}\times\mathcal{B}\times\mathcal{N}}. \end{aligned}$$

We immediately deduce from (6) that for any $v \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}}$

$$R(v) = \gamma\ \|Bv\|^2.$$

We also easily obtain (see Appendix 6.6) that the adjoint B^* of B is provided for any $w \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}\times\mathcal{N}}$ by

$$(B^* w)_{p,q} = \sum_{p'\in\mathcal{N}} w_{p,q,p'} - w_{p-p',q,p'}, \forall (p,q) \in \mathcal{P}\times\mathcal{B}.$$

We can now state the following proposition whose proof is provided in Appendix 6.7.

Proposition 3 *For any $v \in \mathcal{U}^{\mathcal{P}}$, we have*

$$\nabla R(v) = 2\gamma B^* B v.$$

Moreover, $v \to \nabla R(v)$ is Lipschitz continuous with Lipschitz constant $6\sqrt{2}\gamma|\mathcal{N}|$. As a consequence, (10) holds for $l' = 6\sqrt{2}\gamma|\mathcal{N}|$.

3.5 Calculation of $\mathrm{prox}_L^D\,(u')$

Below, we provide closed form expressions that permit to compute $\mathrm{prox}_L^D\,(u')$ for the two data fidelity terms considered in this paper. In the denoising case, H is a identity; for the purpose of image inpainting, H multiply every entry of the image by either 1 or 0 (see (1)).

- **When $D(u) = \lambda\|u - I\|^2$ and $L = l"Id$.**
  In this case, we want to compute

$$\begin{aligned} & \mathrm{prox}_L^{\lambda\|u-I\|^2}\,(u') \\ &= \operatorname*{argmin}_{u\in\mathbb{R}^{\mathcal{P}}} \lambda\|u-I\|^2 + \frac{l"}{2}\|u-u'\|^2, \\ &= \frac{2\lambda}{2\lambda+l"} I + \frac{l"}{2\lambda+l"} u'. \end{aligned}$$

– **When $D(u) = \chi_{\{\|u-I\|^2 \leq \tau\}}(u)$ and $L = l" Id$.**
In that setting, we have

$$\begin{aligned}
&\operatorname{prox}_L^{\chi_{\{\|u-I\|^2\leq\tau\}}}(u') \\
&= \operatorname*{argmin}_{u:\|u-I\|^2\leq\tau} \|u-u'\|_L^2 \\
&= \begin{cases} u' & \text{, if } \|u'-I\|^2 \leq \tau, \\ I + \frac{\sqrt{\tau}}{\|u'-I\|}(u'-I) & \text{, otherwise.} \end{cases}
\end{aligned}$$

– **When $D(u) = \lambda\|Hu - I\|^2$**
In this case,

$$\begin{aligned}
&\operatorname{prox}_L^{\lambda\|H.-I\|^2}(u') \\
&= \operatorname*{argmin}_{u\in\mathbb{R}^{\mathcal{P}}} \lambda\|Hu-I\|^2 + \frac{l"}{2}\|u-u'\|^2, \\
&= u^*,
\end{aligned}$$

where for all $p \in \mathcal{P}$

- if H is defined by (1) for some fixed subset of pixels $\mathcal{P}' \subset \mathcal{P}$

$$u_p^* = \begin{cases} u'_p & \text{, if } p \in \mathcal{P}', \\ \frac{2\lambda I_p + l" u'_p}{2\lambda + l"} & \text{, otherwise.} \end{cases}$$

- if H is a general linear operator (e.g. the subsampling operator used for zooming) then

$$2\lambda H^T(Hu - I) + l"(u - u') = 0,$$

thus u^* can be solved by the conjugate gradient (CG) method from a sparse linear equation as follow

$$Au = b,$$

where $A = \left(2\lambda H^T H + L\right)$ and $b = 2\lambda H^T I + l" u'$. Note that, although we use CG (call Matlab function `pcg`(A,b)) to solve this subproblem, the elementwised max norm of the residual $\max|Au^* - b|$ is around 10^{-13}.

4 Numerical experiments

In order to illustrate the behavior of the model, we consider in this section three applications: image inpainting, zooming and denoising. We evaluate the ability of the method to restore both synthetic and natural images with a particular emphasis on the restoration of thin structures. We compare the performance of the proposed RNLTV model with the performance of state of the art methods. We also illustrate how the weights v are adjusted once the convergence has been reached using the Isotropy Map described in Section 4.1. Finally, Section 4.2 is devoted to the inpainting experiments, Section 4.3 presents zooming results and Section 4.4 contains the experiments on the denoising problem.

Notice that we have used the different data-fidelity terms mentioned in Section 2.1. They all show similar computational cost and convergence properties. For completeness, we summarize the data fidelity term chosen for each experiments in Table 1.

(–5,5)	(–4,5)	...	(4,5)	(5,5)
(–5,4)	(–4,4)	...	(4,4)	(5,4)
⋮	⋮		⋮	⋮
(–5,–4)	(–4,–4)	...	(4,–4)	(5,–4)
(–5,–5)	(–4,–5)	...	(4,–5)	(5,–5)

Fig. 2 The window $\mathcal{B}$ of size 11×11. Every $q \in \mathcal{B}$ has a coordinate.

Table 1 The different data-fidelity terms used in the experiments.

	$D(u)$
inpainting and zooming	$\lambda\|Hu - I\|^2$
denoising (synthetic image)	$\chi_{\{\|u-I\|^2\leq\tau\}}(u)$
denoising (natural image)	$\lambda\|u - I\|^2$

4.1 The Isotropy Map

To represent the distribution of the weights v, we introduce the isotropy map $M(v) \in \mathbb{R}^{\mathcal{P}}$. At every pixel $p \in \mathcal{P}$, the feature $M(v)_p$ represents whether the weights in v^p are distributed isotropically or not. Intuitively, the distribution is not isotropic when the significant weights are concentrated along a thin region of the window $p + \mathcal{B}$. The feature therefore needs to reflect the discrepancy between the projections of v^p onto two axes one of which is aligned with the orientation of v^p, the other being orthogonal. Notice that the discrepancy is maximal when the coordinate system is aligned with the orientation of v^p.

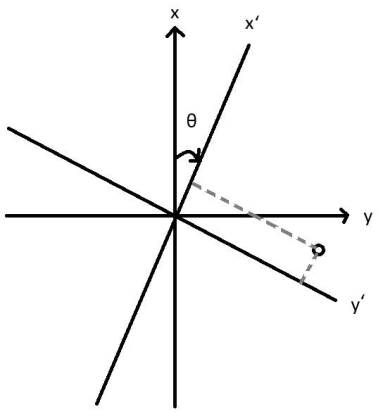

Fig. 3 Coordinate change and projection of the weights on the new coordinate system.

More precisely, given an angular precision parameter $K \in \mathbb{N}$ [3] we define the small angle $\theta = \frac{\pi}{2K}$ as shown in Figure 3. Then we consider all the coordinate (x, y) clock-wise rotated by an angle $m\theta$, for $m \in \mathcal{M}$, with $\mathcal{M} = \{0, 1, 2, \ldots, K-1\}$. Doing so, we get several orthonormal bases denoted by

$$i_m = (\sin(m\theta), \cos(m\theta)),$$
$$j_m = (\cos(m\theta), -\sin(m\theta)).$$

We remind that any pixel $q \in \mathcal{B}$ is represented in the basis (i_m, j_m) by the coordinates

$$(\langle q, i_m \rangle, \langle q, j_m \rangle).$$

Therefore, the average distance to the origin of the projections of v^p onto the axes i_m and j_m are defined by

$$P^x_m = \sum_{q \in \mathcal{B}} \left| v^p_q \langle q, i_m \rangle \right|,$$

$$P^y_m = \sum_{q \in \mathcal{B}} \left| v^p_q \langle q, j_m \rangle \right|.$$

The isotropy map is then defined by

$$M(v)_p = \log\left(\max\left(\left\{ \frac{P^x_m}{P^y_m}, \frac{P^y_m}{P^x_m} \right\}_{m \in \mathcal{M}} \right) \right).$$

Clearly, $M(v)_p$ is non-negative. Also, it is easy to check that the only way to get small $M(v)_p$ is that $P^x_m \sim P^y_m$ for all $m \in \mathcal{M}$. This implies that v^p is somewhat isotropic. On the contrary, if $M(v)_p$ is large, then there exists $m \in \mathcal{M}$ such that P^x_m and P^y_m are dissimilar. From this condition, we conclude that v^p distributes anisotropically. In summary, $M(v)$ can be taken as an indication of the degree of isotropy of v^p, for every pixel $p \in \mathcal{P}$.

[3] In the experiments, we use $K = 90$.

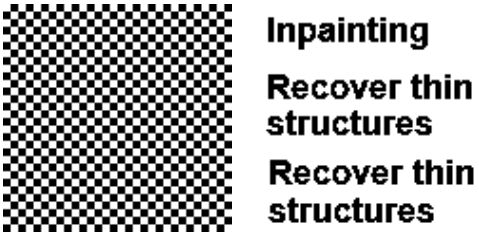

Fig. 4 Inpainting masks with missing data marked in black. Left: checkerboard mask; Right: word mask

4.2 Inpainting

In this section, we use two masks presented in Figure 4, where the black pixels indicate the region to be inpainted. We set values of parameters in RNLTV as follow: $\lambda = 5000$ and $\mu = 0.8$ where we remind that the image graylevels range in $[0, 1]$. The number of iterations[4] is 3000, the radius $r = 5$ and the γ values are in Table 2. We compare the result of the model with some related methods: the MATLAB function inpaintn [23,42], TV inpainting and TV-H^{-1} inpainting [10,40]. All these methods share the particularity of being based on a image regularity criterion. They all share the good property of being easy to adapt to solve other inverse problems.

Table 2 Values of the parameters γ of RNLTV for the restoration of the images in Figure 5, when corrupted by the checkerboard or word mask (see Figure 4). The other parameters are $\mu = 0.8$, $\lambda = 5000$. The restoration results are presented in Figure 6-9.

	checkerboard mask	word mask
	γ	γ
Synthetic	3	2
Boat	3	3
Barbara	4	3
Fingerprint	3	3

4.2.1 Synthetic Image

In order to illustrate the ability of RNLTV to inpaint straight and curved thin structures, we construct a synthetic image with 3 ancient Chinese characters (see Figure 5).

In Figure 6, RNLTV outperforms the other methods. There are two strong strokes in the synthetic

[4] We are aware of the fact that designing a good stopping criterion would permit to save time. However, since the paper presents a new model, we preferred to loose computational time in order to obtain results that truly reflect the behavior of this model.

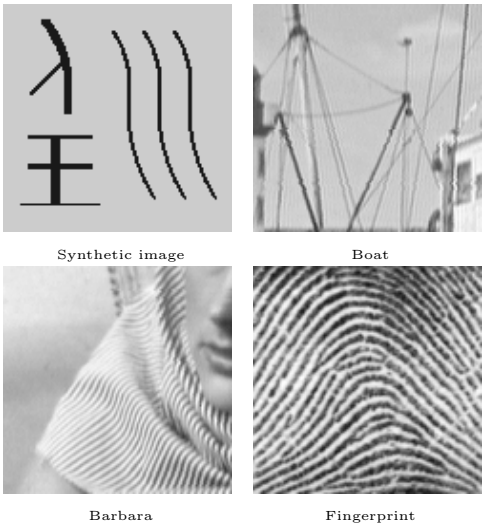

Fig. 5 Test images.

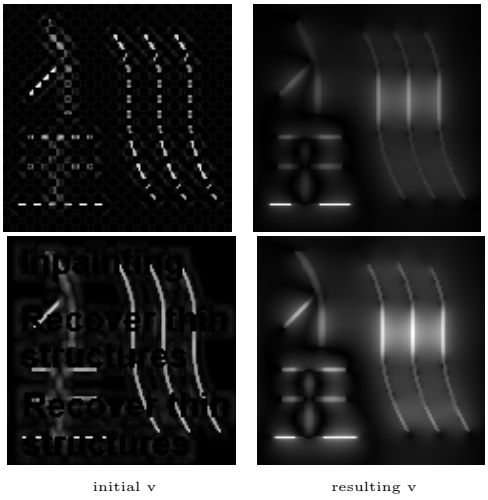

Fig. 7 Plots of the isotropy map $M(v) \in \mathbb{R}^{\mathcal{P}}$ calculated from weights the v in RNLTV. The corresponding restored images u are in Figure 6. Top: for the checkerboard mask; Bottom: for the word mask. Left: initial guesses of the weights computed from the damaged image; Right: weights v after convergence of RNLTV.

image: the arc in the character in the top left corner, and the cross in the middle of the character in the bottom left corner. From the first row of Figure 6, we observe that only RNLTV can connect all the thin edges correctly. The TV regularization propagates the background color into the missing domains (see the 3 curved lines on the right). The method "inpaintn" and TV-H^{-1} cannot clearly inpaint the missing edges in the missing squares on the thin edges. For example, in the middle parts of the 3 curved lines. The method "inpaintn" creates nonexisting white spots, and blurs the rest of the 3 curved lines. The TV-H^{-1} model connects the thin edge with a wrong shape and a wrong intensity. Taking the character on the bottom left corner for example, we can see that TV-H^{-1} fills larger intensity values into some missing squares located on the top horizontal line, which yields lighter color than it is supposed to be, while puts smaller intensity values into some missing squares located on the bottom horizontal line. The RNLTV model connects all the edges correctly except at the end points, see the ends of the 3 curved lines.

In the second row of Figure 6, we display similar restoration results, for inpainting the word mask. The TV regularization connects the disjoint strong objects, but still fails on the thin ones. Again, the algorithm "inpaintn" and TV-H^{-1} cannot recover the images.

In Figure 7, we plot the isotropy map $M(v) \in \mathbb{R}^{\mathcal{P}}$ for the initial and final weights v in RNLTV. This illustrates that RNLTV has the ability to fill the missing regions with reasonable weights. As can be seen in the first column of Figure 7, the initial weights are distributed isotropically both in the regions to be inpainted and the homogeneous regions. However, the regularization term $R(v)$ permits to interpolate "coherent weights" in the missing areas. For example, due to the influence of their neighbors, the inpainted weights located on the lost thin curves are anisotropic and align with their neighbors.

4.2.2 Natural Image

In this section, we inpaint natural images in the regions described by the two masks in Figure 4. The test images are Boat, Barbara and Fingerprint (see Figure 5). They contain isolated thin structures or textures where thin structures are grouped.

In the first row of Figure 8, the TV regularization takes the color of the sky and fills it into the missing areas. In particular, the thin ropes become dashed lines. The methods "inpaintn" and TV-H^{-1} return blurry restoration results. Even worse, the method "inpaintn" creates non-existing white spots around the ropes. The RNLTV model recovers most ropes. In the first row of Figure 8, even the ropes and the mast crossing each other are better restored with RNLTV. Moreover, only RNLTV can recover the mast on the right.

On Figure 9, when inpainting the checkerboard mask, the stripe pattern on the scarf are completely ruined by all the methods except RNLTV. In the second row of Figure 9, we obtain similar

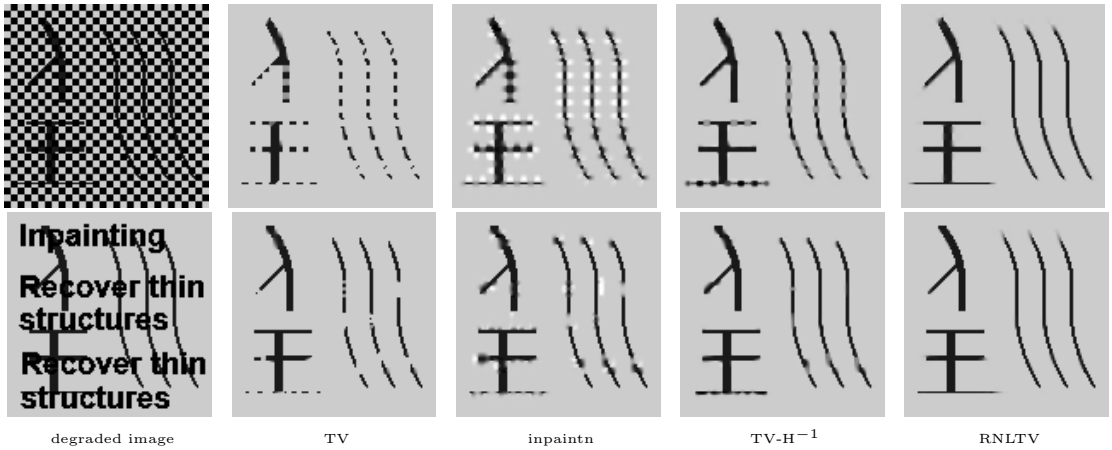

Fig. 6 Results of different methods when restoring the degraded synthetic image with the checkerboard and word masks presented in Figure 4.

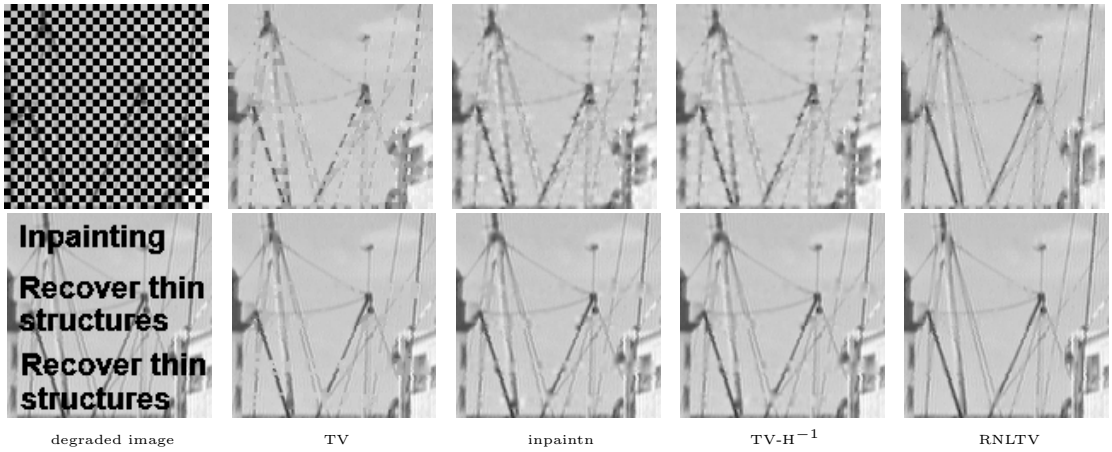

Fig. 8 Restoration results for different methods when inpainting with the checkerboard (Top) and word (Bottom) masks presented in Figure 4.

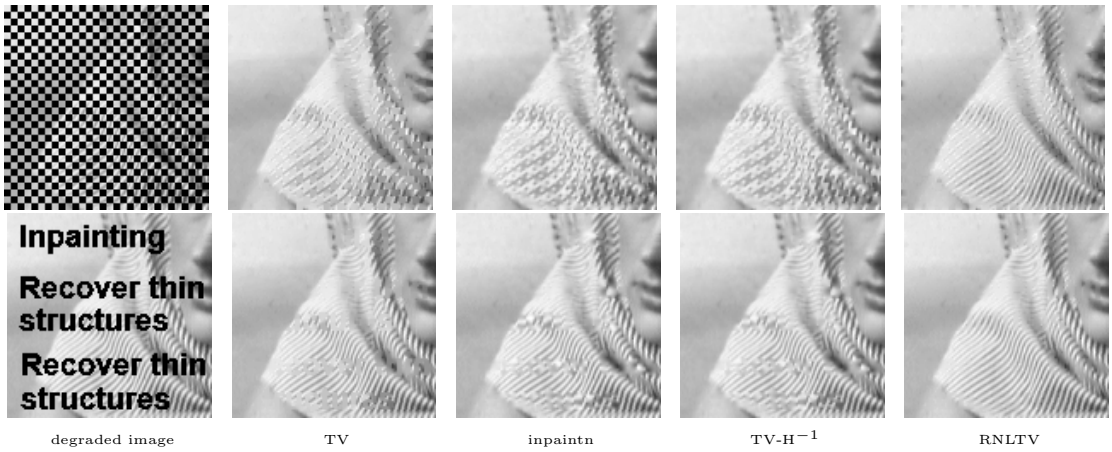

Fig. 9 Restoration results for different methods when inpainting with the checkerboard (Top) and word (Bottom) masks presented in Figure 4.

results. RNLTV still gets clear inpainting results, even on the scarf folds near the face. However, when restoring images corrupted by the checkerboard mask. Of course, since we use periodic boundary condition, we observe some boundary effects with RNLTV. There are many ways to avoid this artifact.

In the first row of Figure 10, the TV regularization and the methods "inpaintn" and TV-H^{-1} create fake lines crossing the friction ridges. Moreover, these results contain many discontinuous friction ridges in the right top corner. The RNLTV model connects all the broken friction ridges and tends to generate well contrasted, smooth ridges. The same comments are true for images on the second row of Figure 10, the results of the methods "inpaintn" and TV-H^{-1} are blurred. In both examples, the TV regularization inpaints the missing areas with wrong colors.

Table 3 PSNR values of different methods when inpainting with the checkerboard and word masks. Sy: Synthetic image; Bo: Boat; Fi: Fingerprint; Ba: Barbara;cb: the checkerboard mask; word: the word mask

	mask	polluted	TV	inpaintn	TV-H^{-1}	RNLTV
Sy	cb	3.38	18.55	19.61	20.01	26.78
	word	7.43	20.80	21.81	23.64	28.97
Bo	cb	5.41	25.53	26.82	26.88	28.60
	word	9.24	15.38	15.26	30.79	33.79
Ba	cb	5.31	24.94	24.50	24.75	28.36
	word	9.24	29.06	28.38	28.64	33.71
Fi	cb	7.08	16.47	19.60	19.25	22.69
	word	11.04	21.08	10.57	24.36	27.73

4.3 Zooming

We compare the zooming results of RNLTV with the results of two methods: the bicubic interpolation and the TV based restoration. In all the experiments, the image data I_d is computed by the Matlab function `imresize` (I_d = `imresize`($\overline{u}$, 0.25)) from an original image $\overline{u}$ with a subsampling of a factor 4, as presented in the first column of Figure 11 and 12. The bicubic interpolation is simply implemented by calling the Matlab function `imresize` again (`imresize`(u_d, 4, 'bicubic')); We set $\lambda = 1000$ and $\mu = 0.6$ in the RNLTV model.

The thin blood vessels are clearly zoomed by RNLTV as presented in the last column of Figure 11. Furthermore, Figure 11 shows a shortcoming of the TV based restoration, that is, it tends to create piecewise constant zones in the homogenous regions which is the tissue around the vessels and reduces the contrast of thin structures. The PSNR values in Table 4 confirm the observation.

Table 4 PSNR values of the different methods for zooming of images downsampled by a factor of 4.

	bicubic	TV	RNLTV
Vessel	34.42	12.69	35.15
Ship	24.58	22.90	24.46

In Table 4, when zooming "Lacournouaille", although the PSNR value of the result of bicubic is slightly higher, from Figure 12, one can observe that RNLTV based zooming leads to a better visual quality whereas the bicubic interpolation gets a more blurry result, see the mast pointing to the one clock direction on the right part of Ship.

4.4 Denoising

In this section, we compare RNLTV to two related regularization methods: the TV denoising (ROF [39]) and the NLTV denoising [25]. For the NLTV, we simply set the γ to 0 in our RNLTV code. The NLTV and RNLTV use identical initial weights for the same problem. The initial weights are obtained by applying (3) in which the image I is replaced by the result of the ROF model (i.e. the image presented in the TV column in the same line as the two non-local methods in Figure 14-16). The initial guess for u is still the degraded image.

4.4.1 Synthetic Image

We consider additive Gaussian white noise with standard deviation $\sigma = 0.12$, 0.14, 0.16. As mentioned in Table 1, we use the data fidelity term $\chi_{\|\|u-I\|^2 \leq \tau}(u)$. We set $\gamma = 0.3$, $\tau = |\mathcal{P}|\sigma^2$ and $\mu = 0.6$ in the RNLTV model. The number of iterations[5] is 600.

In Figure 14, we observe that RNLTV produces sharp edges for all shapes. The contrast is reduced differently in different parts of the thin structures in the ROF results. The black region sometimes "leak" in the background creating some kind of "blur-like" artifact. Also, the NLTV denoising result leave some noise. This is due to the fact that, because of the noise, almost all of the weights distribute in a similar manner and are all isotropic (see the first column of Figure 13). By regularizing the weights v, the RNLTV improves those weights

[5] Again, we just take a very large number of iteration for which we have convergence.

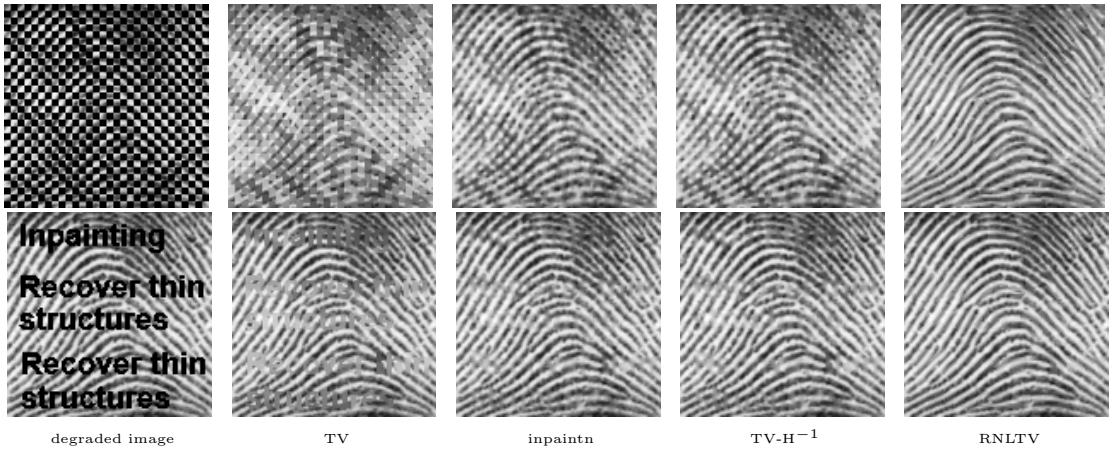

Fig. 10 Restoration results for different methods when inpainting with the checkerboard (Top) and word (Bottom) masks presented in Figure 4.

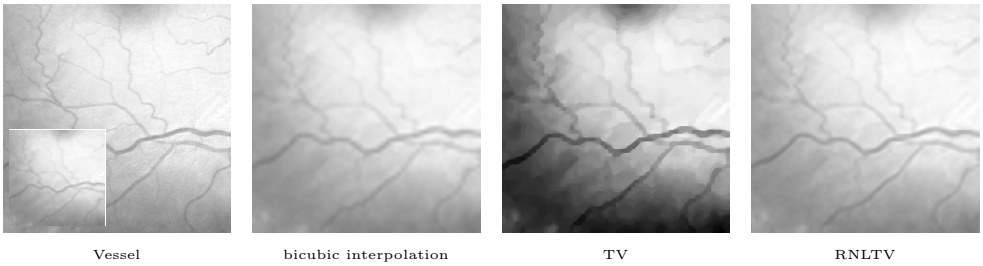

Fig. 11 Image zooming. First column: downsample the original image by a factor of 4.

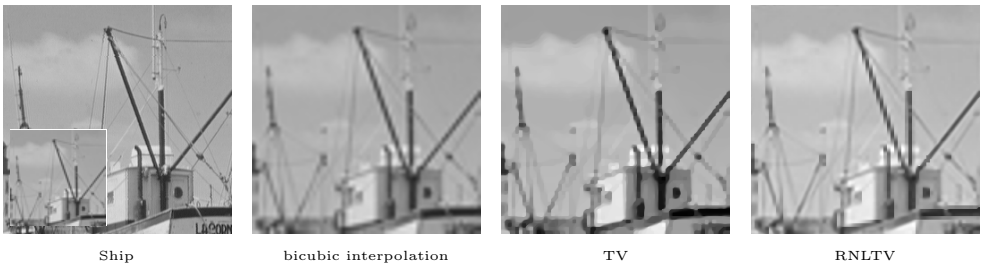

Fig. 12 Image zooming. First column: downsample the original image by a factor of 4.

(see the second column in Figure 13) which leads to a better denoising result.

4.4.2 Natural Image

We consider additive Gaussian white noise with standard deviation $\sigma = 0.04$, 0.06 and 0.08 (We add more noise to Fingerprint, because it is more noise-resistant than the rest of test images). We lower the noise level because the thin structures in the natural test images are less contrasted. We use the results of TV denoising to initialize RNLTV and we use the data fidelity term $\lambda\|u - I\|^2$. The setting of parameters in RNLTV and NLTV is presented in Table 5 and 6 respectively. The number of iterations[6] are both 600. The results are in Figure 15-16.

In all the experiments displayed in Figure 15 (the last row being the more impressive) we observe a similar phenomenon. The TV regularization reduces the contrast or erases the ropes. The RNLTV method clearly better recovers the ropes which are more detectable when compared to the NLTV result. Even the upper part of the mainmast (which is poorly contrasted) are erased in the results of TV and NLTV regularization. The

[6] Again, we just take a very large number of iteration for which we have convergence.

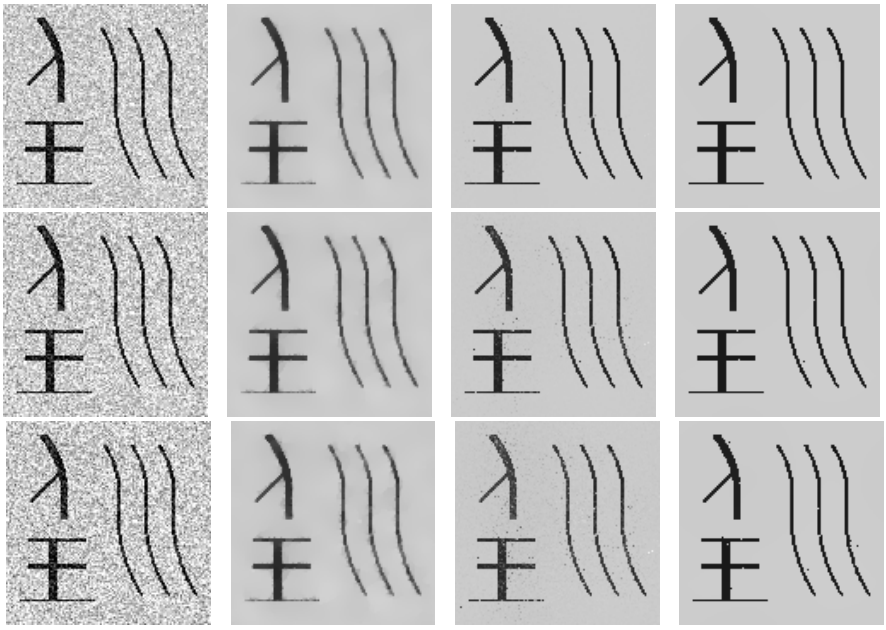

Fig. 14 Denoising results for a additive white Gaussian noise with standard deviation $\sigma = 0.12$ (top row), $\sigma = 0.14$ (middle row) and $\sigma = 0.16$ (bottom row).

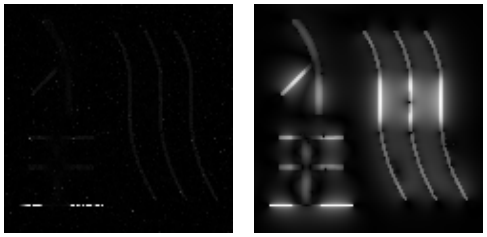

Fig. 13 Plots of the isotropy map $M(v) \in \mathbb{R}^{\mathcal{P}}$ calculated from weights the v in RNLTV. The corresponding denoised images u are in Figure 14. Left: initial guesses of the weights computed from the noisy image (i.e. NLTV); Right: weights v after convergence of RNLTV.

RNLTV model better restores it (see the bottom row of Figure 15).

In Figure 17, the RNLTV model obtains the smoother friction ridges. They are also well contrasted. In the last row of Figure 16, the TV denoising results contains staircasing effects and a large part of the texture of the scarf is erased. Similarly, some gray stripes on the NLTV denoising result become less notable. The RNLTV model recovers the majority of the thin stripes on the scarf to their original shapes.

Table 5 Values of parameters in RNLTV for restoration of the natural images corrupted by white Gaussian noise. μ is fixed to 0.6, σ is the standard deviation of the Gaussian noise.

	σ=0.04		σ=0.06		σ=0.08	
	γ	λ	γ	λ	γ	λ
Boat	0.05	0.5	0.1	0.5	0.2	0.5
Barbara	0.05	0.5	0.1	0.5	0.2	0.5
Fingerprint	0.05	1	0.1	1	0.2	1

Table 6 Values of parameters in NLTV for restoration of the natural images corrupted by white Gaussian noise. μ is fixed to 0.6, σ is the standard deviation of the Gaussian noise.

	σ=0.04	σ=0.06	σ=0.08
	λ	λ	λ
Boat	1	0.5	0.5
Barbara	1	0.5	0.2
	σ=0.1	σ=0.16	σ=0.2
Fingerprint	λ	λ	λ
	0.2	0.1	0.1

5 Conclusion

In this paper, we have studied a RNLTV model to restore the thin structures through the regularization term on weights v, in a NLTV model. In order to implement a PALM solver, we compute the gradients and the Lipschitz moduli of the gradient mapping. The performance of the proposed

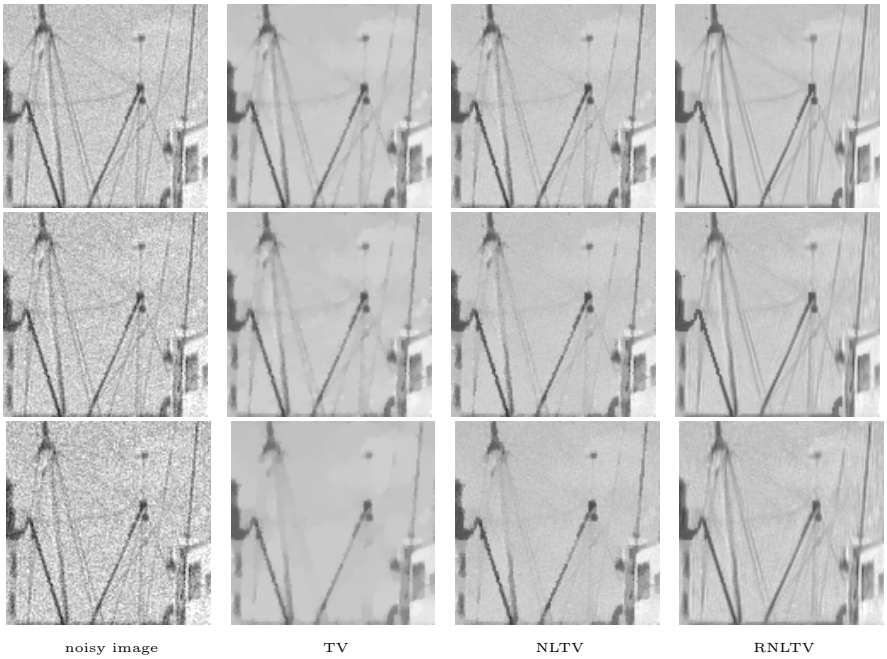

Fig. 15 Denoising results for a additive white Gaussian noise with standard deviation $\sigma = 0.04$ (top row), $\sigma = 0.06$ (middle row) and $\sigma = 0.08$ (bottom row).

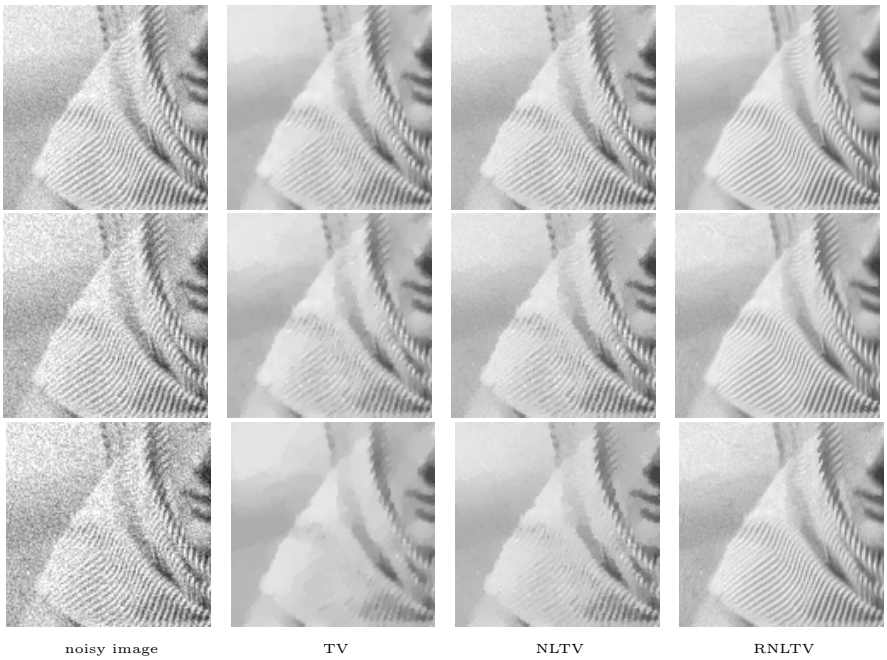

Fig. 16 Denoising results for a additive white Gaussian noise with standard deviation $\sigma = 0.04$ (top row), $\sigma = 0.06$ (middle row) and $\sigma = 0.08$ (bottom row).

Fig. 17 Denoising results for the additive white Gaussian noise with standard deviation $\sigma = 0.1, 0.16, 0.2$ (from top to bottom).

Table 7 PSNR values of different methods for restoration of images corrupted by white Gaussian noise. σ is the standard deviation of the Gaussian noise.

	σ	noisy	TV	NLTV	RNLTV
Syn	0.12	16.54	21.76	31.91	39.06
	0.14	15.20	21.64	27.20	34.04
	0.16	14.04	21.49	23.67	29.65
Boat	0.04	28.02	30.73	31.47	32.25
	0.06	24.50	28.73	29.27	30.26
	0.08	22.00	25.97	27.40	28.86
Barbara	0.04	28.02	29.58	30.64	32.09
	0.06	24.50	26.76	27.68	30.26
	0.08	22.00	23.98	25.48	28.79
Fingerprint	0.1	20.06	21.59	21.63	24.09
	0.16	15.98	19.42	19.58	21.41
	0.2	14.04	18.33	18.34	19.98

method is demonstrated by numerical experiments on synthetic images and natural images. In particular, we see that the regularization of the weights permit to smooth or interpolate them. This leads to better restoration results.

In the future, we plan to improve the ability of the RNLTV model to recover textures, one possible solution is to use shape-adaptive patches [19]. More precisely, to avoid diminishing thin structures, we choose the shape of the patch based on the local geometry, so that the pixels depicting thin structures in a window will take relatively large portion of the overall, thus will not be overlooked, while patches preserve texture areas.

Acknowledgement

François Malgouyres would like to thank Julien Rabin for fruitful discussions on the subject and for teaching him how to efficiently perform the projection on the simplex.

6 Appendices

6.1 Calculation of the adjoint of $\mathbf{D}_v$

Let us first remind that, for any $v \in \mathcal{U}^{\mathcal{P}}$,

$$\mathbf{D}_v : \quad \mathbb{R}^{\mathcal{P}} \longrightarrow \mathbb{R}^{\mathcal{P}\times\mathcal{B}}, \quad u = (u_p)_{p\in\mathcal{P}} \longmapsto \left(\sqrt{v_q^p}\,(u_p - u_{p+q})\right)_{(p,q)\in\mathcal{P}\times\mathcal{B}}.$$

In order to compute its adjoint, we consider $u \in \mathbb{R}^{\mathcal{P}}$, $v \in \mathcal{U}^{\mathcal{P}}$ and $w \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}}$. We have

$$\langle \mathbf{D}_v u, w\rangle = \sum_{q\in\mathcal{B}}\sum_{p\in\mathcal{P}} \sqrt{v_q^p}\,(u_p - u_{p+q})w_{p,q}.$$

Moreover, for any $q \in \mathcal{B}$

$$
\begin{aligned}
&\sum_{p\in\mathcal{P}} \sqrt{v_q^p}\ (u_p - u_{p+q})w_{p,q} \\
&= \sum_{p\in\mathcal{P}} \sqrt{v_q^p} u_p w_{p,q} - \sum_{p\in\mathcal{P}} \sqrt{v_q^p} u_{p+q} w_{p,q}, \\
&= \sum_{p\in\mathcal{P}} \sqrt{v_q^p} u_p w_{p,q} - \sum_{p\in\mathcal{P}} \sqrt{v_q^{p-q}} u_p w_{p-q,q}, \\
&= \sum_{p\in\mathcal{P}} u_p (\sqrt{v_q^p} w_{p,q} - \sqrt{v_q^{p-q}} w_{p-q,q}).
\end{aligned}
$$

Therefore,

$$
\langle \mathbf{D}_v u, w\rangle = \sum_{p\in\mathcal{P}} u_p \left(\sum_{q\in\mathcal{B}} \left(\sqrt{v_q^p} w_{p,q} - \sqrt{v_q^{p-q}} w_{p-q,q} \right) \right),
$$

from which we conclude that, for all $p \in \mathcal{P}$,

$$
(\mathbf{D}_v^* w)_p = \sum_{q\in\mathcal{B}} \left(\sqrt{v_q^p}\ w_{p,q} - \sqrt{v_q^{p-q}}\ w_{p-q,q} \right).
$$

6.2 TV under a max form

We want to prove that for any $u \in \mathbb{R}^{\mathcal{P}}$ and any (fixed) $v \in \mathcal{U}^{\mathcal{P}}$

$$
TV\,(v,u) = \max_{\|w\|_{\infty,2}\leq 1} \langle \mathbf{D}_v u, w\rangle - \frac{\mu}{2}\|w\|^2,
$$

where we remind that, for $w \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}}$, the norm defining the constraint takes the form

$$
\|w\|_{\infty,2} = \max_{p\in\mathcal{P}} \|w_p\|, \text{ for } w_p = (w_{p,q})_{q\in\mathcal{B}} \in \mathbb{R}^{\mathcal{B}}.
$$

First, notice that if for all $p \in \mathcal{P}$ we know $w_p^* \in \mathbb{R}^{\mathcal{B}}$ such that

$$
w_p^* = \text{argmax}_{w\in\mathbb{R}^{\mathcal{B}}:\|w\|\leq 1} \langle \mathbf{D}_v u_p, w\rangle - \frac{\mu}{2}\|w\|^2, \quad (15)
$$

where we denote $\mathbf{D}_v u_p = (\mathbf{D}_v u_{p,q})_{q\in\mathcal{B}} \in \mathbb{R}^{\mathcal{B}}$; we can deduce from the optimality of all its components w_p^* that $w^* = (w_p^*)_{p\in\mathcal{P}} \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}}$ satisfies

$$
w^* = \text{argmax}_{w\in\mathbb{R}^{\mathcal{P}\times\mathcal{B}}:\|w\|_{\infty,2}\leq 1} \langle \mathbf{D}_v u, w\rangle - \frac{\mu}{2}\|w\|^2.
$$

In order to calculate w_p^*, for a given $p \in \mathcal{P}$, we first remark that there exists $\alpha_p^* \geq 0$ such that

$$
w_p^* = \alpha_p^* \mathbf{D}_v u_p.
$$

Then, if we use this expression in (15) we have that

$$
\begin{aligned}
\alpha_p^* &= \text{argmax}_{\alpha\|\mathbf{D}_v u_p\|\leq 1} \alpha\|\mathbf{D}_v u_p\|^2 - \frac{\mu}{2}\alpha^2\|\mathbf{D}_v u_p\|^2, \\
&= \text{argmax}_{\alpha\|\mathbf{D}_v u_p\|\leq 1}\ \alpha(1 - \frac{\mu\alpha}{2}), \\
&= \begin{cases} \frac{1}{\|\mathbf{D}_v u_p\|} & \text{, if } \frac{1}{\mu} > \frac{1}{\|\mathbf{D}_v u_p\|}, \\ \frac{1}{\mu} & \text{, otherwise.} \end{cases}
\end{aligned}
$$

We finally obtain that

$$
w_p^* = \begin{cases} \frac{\mathbf{D}_v u_p}{\mu} & \text{, if } \|(\mathbf{D}_v u)_p\| \leq \mu, \\ \frac{\mathbf{D}_v u_p}{\|(\mathbf{D}_v u)_p\|} & \text{, otherwise.} \end{cases}
$$

This corresponds to the expression of $w^*(u)_p$ in Proposition 1.

If we now use the expression for w_p^* to calculate the objective function, we find that

$$
\begin{aligned}
&\max_{w\in\mathbb{R}^{\mathcal{B}}:\|w\|\leq 1} \langle \mathbf{D}_v u_p, w\rangle - \frac{\mu}{2}\|w\|^2 \\
&= \left\langle \mathbf{D}_v u_p, w_p^* \right\rangle - \frac{\mu}{2}\|w_p^*\|^2 \\
&= \begin{cases} \frac{\|\mathbf{D}_v u_p\|^2}{2\mu} & \text{, if } \|\mathbf{D}_v u_p\| \leq \mu, \\ \|\mathbf{D}_v u_p\| - \frac{\mu}{2} & \text{, otherwise.} \end{cases} \\
&= \Psi_\mu(\|\mathbf{D}_v u_p\|).
\end{aligned}
$$

As a consequence,

$$
\begin{aligned}
&\max_{\|w\|_{\infty,2}\leq 1} \langle \mathbf{D}_v u, w\rangle - \frac{\mu}{2}\|w\|^2 \\
&= \sum_{p\in\mathcal{P}} \left\langle \mathbf{D}_v u_p, w_p^* \right\rangle - \frac{\mu}{2}\|w_p^*\|^2 \\
&= TV(v,u).
\end{aligned}
$$

6.3 Proof of Proposition 1

Let us first remind and adapt to the context of this paper the theorem 1 stated in [34]. This result considers finite dimensional real vector spaces V_1 and V_2, a linear operator $A : V_1 \to V_2$, a parameter $\mu > 0$ and a function

$$
f_\mu(x) = \max_{u\in Q_2} \langle Ax, u\rangle - \frac{\mu}{2}\|u\|^2 \qquad ,\forall x \in V_1
$$

where $Q_2 \subset V_2$ is a closed convex bounded set.

It is stated and proved in [34] that the function f_μ is continuously differentiable at any $x \in V_1$. Moreover, if we denote $u_\mu(x)$ the unique solution of the maximization problem defining f_μ, we have:

$$
\nabla f_\mu(x) = A^* u_\mu(x),
$$

where A^* is the adjoint of A. Moreover, $x \longmapsto f_\mu(x)$ is Lipschitz continuous with constant

$$
\frac{\|A\|_{V_1\to V_2}}{\mu},
$$

where

$$\|A\|_{V_1\to V_2} = \max_{\|x\|\leq 1} \|Ax\|.$$

The Proposition 1 is a straightforward application of this statement to the function

$$TV(v,u) = \max_{\|w\|_{\infty,2}\leq 1} \langle \mathbf{D}_v u, w\rangle - \frac{\mu}{2}\|w\|^2.$$

In order to compute the Lipschitz constant, we need however to find a bound for $\|\mathbf{D}_v\|_{\mathbb{R}^{\mathcal{P}}\to\mathbb{R}^{\mathcal{P}\times\mathcal{B}}}$. In order to compute it, we consider $u\in\mathbb{R}^{\mathcal{P}}$. We have for any $v\in\mathcal{U}^{\mathcal{P}}$

$$\begin{aligned}\|\mathbf{D}_v u\|^2 &= \sum_{p\in\mathcal{P}}\sum_{q\in\mathcal{B}} v_q^p (u_p - u_{p+q})^2\\ &\leq 2\sum_{p\in\mathcal{P}}\sum_{q\in\mathcal{B}} v_q^p\left(u_p^2 + u_{p+q}^2\right)\\ &\leq 2\sum_{p\in\mathcal{P}} u_p^2 + 2\sum_{q\in\mathcal{B}}\sum_{p\in\mathcal{P}} u_{p+q}^2\\ &\leq 2\|u\|^2 + 2|\mathcal{B}|\|u\|^2,\end{aligned}$$

where $|\mathcal{B}|$ denotes the cardinality of $\mathcal{B}$. We then deduce[7] that for any $v\in\mathcal{U}^{\mathcal{P}}$

$$\|\mathbf{D}_v\|_{\mathbb{R}^{\mathcal{P}}\to\mathbb{R}^{\mathcal{P}\times\mathcal{B}}} \leq \sqrt{2}\sqrt{|\mathcal{B}|+1}.$$

Finally, it is standard that (14) is a consequence of the fact that $l" = \frac{\sqrt{2}\sqrt{|\mathcal{B}|+1}}{\mu}$ is a Lipschitz bound for $u\longmapsto \nabla_u TV(v,u)$.

6.4 Calculation of the adjoint of A_u

Let us first remind that for $u\in\mathbb{R}^{\mathcal{P}}$, the operator $\mathbf{A}_u : \mathbb{R}^{\mathcal{P}\times\mathcal{B}} \longrightarrow \mathbb{R}^{\mathcal{P}}$ and is defined for any $v\in\mathcal{U}^{\mathcal{P}}$ and any $p\in\mathcal{P}$ by

$$(\mathbf{A}_u v)_p = \sum_{q\in\mathcal{B}} v_{p,q}(u_p - u_{p+q})^2.$$

Let $u\in\mathbb{R}^{\mathcal{P}}$, $v\in\mathbb{R}^{\mathcal{P}\times\mathcal{B}}$ and $w\in\mathbb{R}^{\mathcal{P}}$, we have

$$\begin{aligned}\langle \mathbf{A}_u v, w\rangle &= \sum_{p\in\mathcal{P}} w_p \sum_{q\in\mathcal{B}} v_{p,q}(u_p - u_{p+q})^2\\ &= \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} v_{p,q}\left(w_p(u_p - u_{p+q})^2\right)\end{aligned}$$

We then deduce that the adjoint $\mathbf{A}_u^*$ of $\mathbf{A}_u$ is given for any $w\in\mathbb{R}^{\mathcal{P}}$ by

$$(\mathbf{A}_u^* w)_{p,q} = w_p(u_p - u_{p+q})^2,\ \forall (p,q)\in\mathcal{P}\times\mathcal{B}.$$

[7] Notice that a similar upper bound is provided in [11] for the usual finite differences.

6.5 Proof of Proposition 2

Considering $v\in\mathcal{U}^{\mathcal{P}}$, $u\in\mathbb{R}^{\mathcal{P}}$ and a small variation $h\in\mathbb{R}^{\mathcal{P}\times\mathcal{B}}$, we denote $\mathcal{P}_1 = \{p\in\mathcal{P} | (\mathbf{A}_u v)_p \geq \frac{\mu}{2}\}$ and $\mathcal{P}_2 = \mathcal{P}\setminus\mathcal{P}_1$. For h small enough, we have

$$\begin{aligned}TV(v+h,u) = &\sum_{p\in\mathcal{P}_1} \Psi_\mu\left(\sqrt{(\mathbf{A}_u v)_p + (\mathbf{A}_u h)_p}\right)\\ &+ \sum_{p\in\mathcal{P}_2} \frac{(\mathbf{A}_u v)_p + (\mathbf{A}_u h)_p}{2\mu}.\end{aligned}$$

Moreover,

$$\begin{aligned}&\sum_{p\in\mathcal{P}_1} \Psi_\mu\left(\sqrt{(\mathbf{A}_u v)_p + (\mathbf{A}_u h)_p}\right)\\ &= \sum_{p\in\mathcal{P}_1} \Psi_\mu\left(\sqrt{(\mathbf{A}_u v)_p} + \frac{(\mathbf{A}_u h)_p}{2\sqrt{(\mathbf{A}_u v)_p}} + o\left(|(\mathbf{A}_u h)_p|\right)\right)\\ &= \sum_{p\in\mathcal{P}_1} \Psi_\mu\left(\sqrt{(\mathbf{A}_u v)_p}\right)\\ &\quad + \Psi_\mu'\left(\sqrt{(\mathbf{A}_u v)_p}\right)\frac{(\mathbf{A}_u h)_p}{2\sqrt{(\mathbf{A}_u v)_p}} + o\left(|(\mathbf{A}_u h)_p|\right)\end{aligned}$$

Denoting for all $p\in\mathcal{P}_1$

$$(u^*(v))_p = \frac{\Psi_\mu'\left(\sqrt{(\mathbf{A}_u v)_p}\right)}{2\sqrt{(\mathbf{A}_u v)_p}}$$

and for all $p\in\mathcal{P}_2$

$$(u^*(v))_p = \frac{1}{2\mu}$$

and using the simple closed form expression for the derivative Ψ_μ', we get for all $p\in\mathcal{P}$

$$(u^*(v))_p = \begin{cases}\frac{1}{2\sqrt{(\mathbf{A}_u v)_p}} & \text{, if } \sqrt{(\mathbf{A}_u v)_p}\geq\mu,\\ \frac{1}{2\mu} & \text{, if } \mu\geq\sqrt{(\mathbf{A}_u v)_p}\geq 0.\end{cases}$$

Using this notation in the previous calculations we obtain that

$$TV(v+h,u) = TV(v,u) + \langle u^*(v), \mathbf{A}_u h\rangle + o(\|h\|).$$

We finally conclude that

$$\nabla_v TV(v,u) = \mathbf{A}_u^* u^*(v),$$

This proves the first part of Proposition 2.

Let us now show that $v\longmapsto TV(v,u)$ is concave over $\mathbb{R}_+^{\mathcal{P}\times\mathcal{B}}$. In order to do so, we rewrite the latter formula under the form $(u^*(v))_p = \varphi((\mathbf{A}_u v)_p)$ where the function φ is defined for all $t\geq 0$ by

$$\varphi(t) = \begin{cases}\frac{1}{2\sqrt{t}} & \text{, if } \sqrt{t}\geq\mu,\\ \frac{1}{2\mu} & \text{, if } \mu\geq\sqrt{t}\geq 0.\end{cases}$$

Notice that the function φ is non-increasing and therefore

$$(\varphi(t_1) - \varphi(t_2))(t_1 - t_2) \leq 0, \ \forall (t_1, t_2) \in \mathbb{R}^2. \quad (16)$$

We now consider v and $v' \in \mathbb{R}_+^{\mathcal{P}\times\mathcal{B}}$ and $u \in \mathbb{R}^{\mathcal{P}}$. Using Taylor's Theorem, we know there exists $t \in [0,1]$ such that $v" = tv' + (1-t)v$ satisfies

$$\begin{aligned}
&TV(v', u) - TV(v, u) - \langle \nabla_v TV(v, u), v' - v \rangle \\
&= \langle \nabla_v TV(v", u) - \nabla_v TV(v, u), v' - v \rangle \\
&= \langle u^*(v") - u^*(v), \mathbf{A}_u(v' - v) \rangle \\
&= \sum_{p\in\mathcal{P}} \left(\varphi(\mathbf{A}_u v"_p) - \varphi(\mathbf{A}_u v_p)\right) \left(\mathbf{A}_u v'_p - \mathbf{A}_u v_p\right).
\end{aligned}$$

However, for any $p \in \mathcal{P}$, $\mathbf{A}_u v"_p - \mathbf{A}_u v_p = t(\mathbf{A}_u v'_p - \mathbf{A}_u v_p)$ and $\mathbf{A}_u v"_p - \mathbf{A}_u v_p$ and $\mathbf{A}_u v'_p - \mathbf{A}_u v_p$ have the same sign. Using (16), we then get

$$\left(\varphi(\mathbf{A}_u v"_p) - \varphi(\mathbf{A}_u v_p)\right) \left(\mathbf{A}_u v'_p - \mathbf{A}_u v_p\right) \leq 0$$

and finally

$$TV(v', u) - TV(v, u) - \langle \nabla_v TV(v, u), v' - v \rangle \leq 0.$$

This concludes the proof.

6.6 Calculation of the adjoint of B

Let us remind that $B : \mathbb{R}^{\mathcal{P}\times\mathcal{B}} \longrightarrow \mathbb{R}^{\mathcal{P}\times\mathcal{B}\times\mathcal{N}}$ and that for any $v = (v_{p,q})_{(p,q)\in\mathcal{P}\times\mathcal{B}}$ its image Bv is defined at any $(p, q, p') \in \mathcal{P} \times \mathcal{B} \times \mathcal{N}$ by

$$(Bv)_{p,q,q'} = v_{p,q} - v_{p+p',q}$$

We consider $v \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}}$ and $w \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}\times\mathcal{N}}$, we have

$$\begin{aligned}
&\langle Bv, w \rangle \\
&= \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} \sum_{p'\in\mathcal{N}} (v_{p,q} - v_{p+p',q}) w_{p,q,p'}, \\
&= \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} \sum_{p'\in\mathcal{N}} v_{p,q} w_{p,q,p'} \\
&\quad - \sum_{p'\in\mathcal{N}} \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} v_{p+p',q} w_{p,q,p'}, \\
&= \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} \sum_{p'\in\mathcal{N}} w_{p,q,p'} \\
&\quad - \sum_{p'\in\mathcal{N}} \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} v_{p,q} w_{p-p',q,p'}, \\
&= \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} v_{p,q} \left(\sum_{p'\in\mathcal{N}} w_{p,q,p'} - w_{p-p',q,p'} \right).
\end{aligned}$$

We then deduce that the adjoint B^* of B is provided for any $w \in \mathbb{R}^{\mathcal{P}\times\mathcal{B}\times\mathcal{N}}$ by

$$(B^* w)_{p,q} = \sum_{p'\in\mathcal{N}} w_{p,q,p'} - w_{p-p',q,p'},$$

whatever $(p, q) \in \mathcal{P} \times \mathcal{B}$.

6.7 Proof of Proposition 3

For any $v \in \mathcal{U}^{\mathcal{P}}$, given the expression

$$R(v) = \gamma \ \|Bv\|^2$$

we immediately have

$$\nabla R(v) = 2\gamma B^* B v.$$

We only need to calculate the Lipschitz constant l' provided in Proposition 3. In order to do so, we consider v and $v' \in \mathcal{U}^{\mathcal{P}}$ and denote $w = v' - v$. We have

$$\|\nabla R(v') - \nabla R(v)\|^2 = 4\gamma^2 \|B^* B w\|^2.$$

Moreover, using the formula for B^*, we get

$$\|B^* B w\|^2 = \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} \left| \sum_{p'\in\mathcal{N}} \left((Bw)_{p,q,p'} - (Bw)_{p-p',q,p'}\right) \right|^2.$$

The term inside the absolute value can be rewritten, using the definition of B, under the form

$$\begin{aligned}
&\sum_{p'\in\mathcal{N}} \left((Bw)_{p,q,p'} - (Bw)_{p-p',q,p'}\right) \\
&= \sum_{p'\in\mathcal{N}} (2w_{p,q} - w_{p+p',q} - w_{p-p',q}), \\
&= 2|\mathcal{N}| w_{p,q} - \sum_{p'\in\mathcal{N}} w_{p+p',q} - \sum_{p'\in\mathcal{N}} w_{p-p',q}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\|B^* B w\|^2 \\
&\leq \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} 3 \left(4|\mathcal{N}|^2 w_{p,q}^2 + \left(\sum_{p'\in\mathcal{N}} w_{p+p',q} \right)^2 \right. \\
&\qquad \left. + \left(\sum_{p'\in\mathcal{N}} w_{p-p',q} \right)^2 \right) \\
&\leq 3 \left(4|\mathcal{N}|^2 \|w\|^2 + |\mathcal{N}| \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} \sum_{p'\in\mathcal{N}} w_{p+p',q}^2 \right. \\
&\qquad \left. + |\mathcal{N}| \sum_{(p,q)\in\mathcal{P}\times\mathcal{B}} \sum_{p'\in\mathcal{N}} w_{p-p',q}^2 \right) \\
&\leq 3 \times 6 |\mathcal{N}|^2 \|w\|^2.
\end{aligned}$$

Finally,

$$\|\nabla R(v') - \nabla R(v)\|^2 \leq 2 \times 6^2\gamma^2|\mathcal{N}|^2\|v' - v\|^2.$$

We then conclude that $v \longmapsto \nabla R(v)$ is Lipschitz with Lipschitz constant $6\sqrt{2}\gamma|\mathcal{N}|$. This concludes the proof.

References

1. Arias, P., Caselles, V., Sapiro, G.: A variational framework for non-local image inpainting. In: Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 345–358. Springer (2009)
2. Aujol, J.F., Dossal, C.: Stability of over-relaxations for the forward-backward algorithm, application to fista. SIAM J. Optim. **25**(4), 2408–2433 (2015)
3. Aujol, J.F., Ladjal, S., Masnou, S.: Exemplar-based inpainting from a variational point of view. SIAM J. Math. Anal. **42**(3), 1246–1285 (2010)
4. Ballester, C., Bertalmio, M., Caselles, V., Sapiro, G., Verdera, J.: Filling-in by joint interpolation of vector fields and gray levels. IEEE Trans. Image Process. **10**(8), 1200–1211 (2001)
5. Bauschke, H.H., Burachik, R., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H.: Fixed-point algorithms for inverse problems in science and engineering, vol. 49. Springer Science & Business Media (2011)
6. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program., Serie A **146**(1-2), 459–494 (2014)
7. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition, vol. 2, pp. 60–65. IEEE (2005)
8. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Model. Simul. **4**(2), 490–530 (2005)
9. Buades, A., Coll, B., Morel, J.M.: Non-Local Means Denoising. Image Processing On Line **1** (2011)
10. Burger, M., He, L., Schönlieb, C.B.: Cahn-hilliard inpainting and a generalization for grayvalue images. SIAM J. Imaging Sci. **2**(4), 1129–1167 (2009)
11. Chambolle, A.: An algorithm for total variation minimization and applications. Math. Imaging Vision **20**(1-2), 89–97 (2004)
12. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vision **40**(1), 120–145 (2011)
13. Chan, T.F., Shen, J.: Nontexture inpainting by curvature-driven diffusions. J. Vis. Commun. Image Represent **12**(4), 436–449 (2001)
14. Chen, A., Bertozzi, A.L., Ashby, P.D., Getreuer, P., Lou, Y.: Enhancement and recovery in atomic force microscopy images. In: Excursions in Harmonic Analysis, Volume 2, pp. 311–332. Springer (2013)
15. Chen, C., Chan, R.H., Ma, S., Yang, J.: Inertial proximal ADMM for linearly constrained separable convex optimization. SIAM J. Imaging Sci. **8**(4), 2239–2267 (2015)
16. Cho, M., Cai, J.F., Liu, S., Eldar, Y., Xu, W.: Filling-in by joint interpolation of vector fields and gray levels. In: 41st IEEE International Conference on Acoustics, Speech and Signal Processing (2016)
17. Chouzenoux, E., Pesquet, J.C., Repetti, A.: A block coordinate variable metric forward-backward algorithm. J. Global Optim. pp. 1–29 (2016)
18. Condat, L.: Fast Projection onto the Simplex and the ℓ_1 Ball. Math. Program. **158**(1), 575–585 (2016)
19. Deledalle, C.A., Duval, V., Salmon, J.: Non-local methods with shape-adaptive patches (NLM-SAP). J. Math. Imaging Vision **43**(2), 103–120 (2012)
20. Duval, V., Aujol, J.F., Gousseau, Y.: On the parameter choice for the non-local means. CMLA Preprint (2010)
21. Elmoataz, A., Lezoray, O. and Bougleux, S.: Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. IEEE Trans. Image Process. **17**(2), 1047-1060 (2008)
22. Facciolo, G., Arias, P., Caselles, V., Sapiro, G.: Exemplar-based interpolation of sparsely sampled images. In: Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 331–344. Springer (2009)
23. Garcia, D.: Robust smoothing of gridded data in one and higher dimensions with missing values. Comput. Statist. Data Anal. **54**(4), 1167–1178 (2010)
24. Gilboa, G., Osher, S.: Nonlocal linear image regularization and supervised segmentation. Multiscale Model. Simul. **6**(2), 595–630 (2007)
25. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Model. Simul. **7**(3), 1005–1028 (2008)
26. Hu, H., Froment, J.: Nonlocal total variation for image denoising. In: Photonics and Optoelectronics (SOPO), 2012 Symposium on, pp. 1–4. IEEE (2012)
27. Jung, M., Vese, L.A.: Nonlocal variational image deblurring models in the presence of gaussian or impulse noise. In: Scale Space and Variational Methods in Computer Vision, pp. 401–412. Springer (2009)
28. Kanizsa, G.: Organization in vision: Essays on Gestalt perception. Praeger Publishers (1979)
29. Lederman, C.: Two dimensional image denoising using partial differential equations with a constraint on high dimensional domains. Technical Report, Department of Mathematics, University of California, Los Angeles (2011)
30. Levina, E., Bickel, P.J.: Texture synthesis and nonparametric resampling of random fields. Ann. Statist. pp. 1751–1773 (2006)
31. Lintner, S., Malgouyres, F.: Solving a variational image restoration model which involves L^∞ constraints. Inverse Problems **20**(3), 815 (2004)
32. Lou, Y., Zhang, X., Osher, S., Bertozzi, A.: Image recovery via nonlocal operators. J. Sci. Comput. **42**(2), 185–197 (2010)
33. Mumford, D., Nitzberg, M., Shiota, T.: Filtering, segmentation and depth. Lecture Notes in Comput. Sci. **662** (1993)
34. Nesterov, Y.: Smooth minimization of non-smoth functions. Math. Program., Serie A **103**(1), 127–152 (2005)
35. Ouyang, Y., Chen, Y., Lan, G., Pasiliao Jr, E.: An accelerated linearized alternating direction method of multipliers. SIAM J. Imaging Sci. **8**(1), 644–681 (2015)
36. Peyré, G.: Sparse modeling of textures. J. Math. Imaging Vision **34**(1), 17–31 (2009)
37. Peyré, G., Bougleux, S., Cohen, L.: Non-local regularization of inverse problems. In: Computer Vision–ECCV 2008, pp. 57–68. Springer (2008)
38. Peyré, G., Bougleux, S., Cohen, L.D.: Non-local regularization of inverse problems. Inverse Probl. Imaging **5**(2), 511–530 (2011)

39. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**(1), 259–268 (1992)
40. Schönlieb, C.b., Bertozzi, A.: Unconditionally stable schemes for higher order inpainting. Commun. Math. Sci. **9**(2), 413–457 (2011)
41. Shen, J., Chan, T.F.: Mathematical models for local nontexture inpaintings. SIAM J. Appl. Math. **62**(3), 1019–1043 (2002)
42. Wang, G., Garcia, D., Liu, Y., De Jeu, R., Dolman, A.J.: A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. Environ. Modell. Softw. **30**, 139–142 (2012)
43. Wei, L.Y., Levoy, M.: Fast texture synthesis using tree-structured vector quantization. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp. 479–488. ACM Press/Addison-Wesley Publishing Co. (2000)
44. Weiss, P., Blanc-Feraud, L., Aubert, G.: Efficient schemes for total variation minimization under constraints in image processing. SIAM J. Sci. Comput. **31**(3), 2047–2080 (2009)
45. Weissman, T., Ordentlich, E., Seroussi, G., Verdú, S. and Weinberger, M.J.: Universal discrete denoising: Known channel. IEEE Trans. Image Process. **51**(1), 5–28 (2005)
46. Yaroslavsky, L.P., Yaroslavskij, L.: Digital picture processing. an introduction. Springer-Verlag **1** (1985)
47. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J. Imaging Sci. **3**(3), 253–276 (2010)