

EDGE-PRESERVING RECONSTRUCTION WITH CONTOUR-LINE SMOOTHING AND NON-QUADRATIC DATA-FIDELITY

MARC C. ROBINI AND YUEMIN ZHU

CREATIS (CNRS research unit UMR5220 and INSERM research unit U1044)
INSA-Lyon, 69621 Villeurbanne Cedex, France

JIANHUA LUO

College of Life Science and Technology
Shanghai Jiao Tong University, Shanghai 200240, China

(Communicated by Xuecheng Tai)

ABSTRACT. The standard approach for image reconstruction is to stabilize the problem by including an edge-preserving roughness penalty in addition to faithfulness to the data. However, this methodology produces noisy object boundaries and creates a staircase effect. State-of-the-art methods to correct these undesirable effects either have weak convergence guarantees or are limited to specific situations; furthermore, most of them use a quadratic data-fidelity term. In this paper, we propose a simple alternative regularization model to improve contour regularity and to reduce the staircase effect—our model incorporates the smoothness of the edge field in an implicit way by adding a simple penalty term defined in the wavelet domain. We also derive an efficient half-quadratic algorithm to solve the resulting optimization problem, including the case when the data-fidelity term is not quadratic and the cost function is not convex. Our approach either extends or supplements existing methods and offers strong convergence guarantees. Numerical experiments show that it outperforms first-order total variation regularization as well as state-of-the-art second-order regularization techniques.

1. Introduction. We consider the classical inverse problem of recovering a piecewise-smooth image $x^* \in \mathbb{R}^K$ from some data $d \in \mathbb{R}^{K'}$ ($K' \leq K$) of the form

$$(1) \quad d = \chi(\mathcal{H}x^* + \epsilon_\eta),$$

where \mathcal{H} is a linear map from \mathbb{R}^K to $\mathbb{R}^{K'}$ which models the acquisition process, $\epsilon_\eta \in \mathbb{R}^{K'}$ is an unknown realization of a random vector η , and the function $\chi : \mathbb{R}^{K'} \rightarrow \mathbb{R}^{K'}$ represents possible additional corruption by impulse noise. (We call this problem *image denoising* when $\mathcal{H} = \text{id}_{\mathbb{R}^K}$.) Throughout this paper, we make no distinction between (i) a real-valued image with n pixels and its vector representation in \mathbb{R}^n , (ii) a vector of \mathbb{R}^n and its $n \times 1$ column-matrix representation, and (iii) a linear map from \mathbb{R}^n to \mathbb{R}^m and its $m \times n$ canonical matrix. We denote by $\mathcal{L}_{n,m}(\mathbb{R})$ the set of linear maps from \mathbb{R}^n to \mathbb{R}^m .

2010 *Mathematics Subject Classification.* 94A08, 49N45, 68U10, 65K10, 49M20.

Key words and phrases. Image restoration, linear inverse problems, edge-preserving regularization, half-quadratic algorithms, non-convex optimization.

This work was partly supported by the French ANR (ANR-09-BLAN-0372-01) and by the Chinese NSFC (30911130364).

A common estimate of the original image x^* is defined as any global minimum of a cost function $U : \mathbb{R}^K \rightarrow \mathbb{R}$ which combines a data-fidelity term Φ_d with a stabilization term Ψ weighted by a parameter $\lambda > 0$:

$$(2) \quad U(x) = \Phi_d(x) + \lambda\Psi(x).$$

The original justification for this choice lies within the regularization framework [38] and its Bayesian interpretation [17].

The data-fidelity term accounts for the noise characteristics; it is generally defined by

$$(3) \quad \Phi_d(x) = \sum_{k \in \llbracket 1, K' \rrbracket} \phi_k(|(\mathcal{H}x - d)_k|),$$

where the ϕ_k 's are increasing in \mathbb{R}_+ ($\llbracket 1, K' \rrbracket$ is a shorthand notation for the set $\{1, \dots, K'\}$). In the Bayesian setting, this form is obtained in the absence of impulse noise and under the following assumptions: (i) the noise components $\eta_1, \dots, \eta_{K'}$ are independent, and (ii) the density of each η_k is proportional to $\exp(-\phi_k(|t|))$. The most commonly used examples are the squared ℓ_2 -norm of the residual, defined by $\phi_k(t) = t^2$, and the ℓ_1 -norm of the residual, defined by $\phi_k(t) = t$. (The former corresponds to i.i.d. Gaussian random variables, and the latter corresponds to i.i.d. Laplace random variables.) The ℓ_1 -norm also allows to deal with impulse noise [41]. In this case, a possible refinement is to detect the data pixels corrupted by impulse noise and to consider the data to be missing at these outlier pixel locations [7, 8, 34]; the data-fidelity term is then of the same form as (3), but with a summation over the set of non-outlier pixels.

The regularization functional Ψ is intended to promote the formation of smooth regions separated by edges; it is of the form

$$(4) \quad \Psi(x) = \sum_{l \in \llbracket 1, L \rrbracket} \psi_l(\|\mathcal{R}_l x\|_2),$$

where $\|\cdot\|_2$ is the ℓ_2 -norm, $\mathcal{R}_l \in \mathcal{L}_{K, \rho_l}(\mathbb{R})$ ($\rho_l \in \mathbb{N}^*$), and the functions ψ_l are increasing in \mathbb{R}_+ . Concrete examples can be found in [22, 33, 11, 16, 46, 41, 47]. Usually, all the ψ_l 's are equal to a same function ψ , and $\{\mathcal{R}_l : l \in \llbracket 1, L \rrbracket\}$ is a set of first-order difference operators ($\rho_l = 1$) or a discrete approximation to the gradient ($\rho_l = 2$). The ψ_l 's are called *potential functions* (PFs) in the Bayesian framework. They are often classified in two categories: convex PFs, which ensure the convexity of U and reduce smoothing in the vicinity of discontinuities [11], and non-convex PFs, which yield sharper edges [42] at the expense of increased optimization difficulty.

The formulation of the image reconstruction problem as the minimization of a cost function of the form (2)–(4) has proven effective, and yet it has two limitations:

1) *Noisy contour lines*. Regularization functionals of the type of (4) do not embed prior knowledge on the geometry of edges. They tend to produce noisy object boundaries which are not faithful to the original image. Early attempts to correct this effect were based on the introduction of an explicit edge-process—either boolean [23, 3] or continuous [6, 29]—to model the mutual dependence between neighboring discontinuities. However, boolean edge-processes drastically increase computational complexity (not to mention the problem of correctly specifying the penalties of the various edge configurations), and continuous edge-processes give disappointing results.

2) *Staircase effect.* A basic requirement for edge preservation is that $\psi'(t)/t$ goes to zero as $t \rightarrow +\infty$ (see [11]), which explains why convex PFs are typically nearly affine beyond a neighborhood of the origin. In fact, convex edge-preserving PFs are often similar to the identity on \mathbb{R}_+ , but with zero derivative at the origin to make the regularization term Ψ differentiable—a well-known example is

$$(5) \quad \psi_{\text{id}}^\delta(t) = \sqrt{\delta^2 + t^2} - \delta,$$

where $\delta > 0$ is small compared to the width of the range of the original image. If ψ is the identity on \mathbb{R}_+ , $\Psi(x)$ is the discrete *total variation* (TV) of x . Despite its popularity, TV regularization produces blocky images (see, for instance, [19, 10]), and so do PFs close to the identity. This phenomenon is called the *staircase effect* because ramps (affine regions) tend to turn into stairs (piecewise-constant regions); it is most pronounced for bounded PFs [22].

State-of-the-art methods for edge continuation and/or reduction of the staircase effect use second-order difference operators [10, 35, 1, 34, 49, 32, 28, 53]. These methods are reviewed in Section 2; they lack strong convergence guarantees, and, except for [34], they are restricted to quadratic data-fidelity and/or image denoising. Our contributions are the following:

1) We propose a simple alternative regularization model to encourage the formation of smooth contours lines (while preserving edges) and to reduce the staircase effect. Our model is obtained by adding constraints in a multiresolution space: the solutions are the global minima of an augmented functional $V : \mathbb{R}^K \rightarrow \mathbb{R}$ of the form

$$(6) \quad V(x) = U(x) + \tilde{\lambda} \tilde{\Psi}(\mathcal{T}x),$$

where \mathcal{T} is a directional multiresolution transform, $\tilde{\Psi}$ is a smoothness penalty term similar to (4) which operates on the detail coefficients, and $\tilde{\lambda} > 0$ weights the influence of $\tilde{\Psi}$. This technique incorporates the smoothness features of the edge field implicitly and has limited impact on optimization complexity.

2) We derive an efficient deterministic relaxation algorithm to minimize V under differentiability and coercivity assumptions. Our algorithm can deal with non-quadratic data-fidelity and has the following properties: (i) it finds the global minimum when V is strictly convex, (ii) it converges from any initialization when the minima are isolated, and (iii) it gets arbitrarily close to the set of local minima in any situation. The proposed optimization scheme has two further advantages: it allows to enforce inequality constraints on the pixel values, and it allows to use continuation sequences to gradually reveal the complexity of minimizing V (this latter refinement improves performance when V is not convex and/or approximates a non-differentiable cost function).

The price of the convergence guarantees of our deterministic relaxation algorithm is that the PFs must be smooth at zero and coercive (a function ψ defined on \mathbb{R}_+ is said to be smooth at zero if $\psi'(0) = 0$, and it is called coercive if $\lim_{t \rightarrow +\infty} \psi(t) = +\infty$). These conditions preclude PFs of the form $\psi(t) = t^p$ with $p \in (0, 1]$ [47] and bounded PFs such as $\psi(t) = t^p / (\delta^p + t^p)$ with $p \in [1, 2]$ [24, 22]. However, they allow standard choices such as, for instance, ψ_{id}^δ (see (5)), $\psi(t) = \ln(\cosh(t/\delta))$ [25], and the non-convex PF

$$(7) \quad \psi_{\text{HL}}^\delta(t) = \ln(1 + (t/\delta)^2)$$

originally proposed by Hebert and Leahy [26]. In fact, the limitations imposed by our extra conditions are of no practical importance; indeed, any PF that is not

smooth at zero can be closely approximated by a smooth-at-zero PF, and coercive non-convex PFs such as ψ_{HL}^δ perform as well as bounded PFs in terms of edge preservation [42, Theorem 3.1].

This paper is organized as follows. We review previous work related to ours in Section 2. In Section 3, we discuss our approach to the issue of edge continuation. Section 4 is devoted to the description of the optimization algorithm along with its convergence properties (the proofs are relegated to Appendix A). Experimental results are presented in Section 5, followed by concluding remarks.

2. Related work.

2.1. Methods restricted to quadratic data-fidelity. Chouzenoux et al. [13] proposed a subspace optimization approach for minimizing cost functions of the form

$$(8) \quad U(x) = \|\mathcal{H}x - d\|_2^2 + \lambda \sum_l \psi(\|\mathcal{R}_l x\|_2),$$

where $\mathcal{R}_l \in \mathcal{L}_{K,\rho}(\mathbb{R})$ for some fixed $\rho \in \mathbb{N}^*$ and ψ is smooth at zero and coercive. They showed that, under assumptions similar to ours, the sequence $(x^{(n)})_n$ of iterates generated by their algorithm is such that (i) $(U(x^{(n)}))_n$ is non-increasing, and (ii) $\lim_{n \rightarrow \infty} \|\nabla U(x^{(n)})\| = 0$ (as we will see in Section 4.2, these convergence guarantees are weak compared to those derived for our algorithm). Cost functions of the above form were also studied by Nikolova et al. [43]. They developed a continuation approach to handle the situation where ψ is not smooth at zero and not convex, but they did not investigate the theoretical convergence properties of their algorithms. In [27], Hou et al. considered the special case where $\psi(t) = t^p$ with $p \in (0, 1)$ in order to promote sparsity of the gradient at different resolutions (the set $\{\mathcal{R}_l\}_l$ consists of multi-resolution derivative filters). They used the multi-stage convex relaxation method proposed in [52], whose only convergence guarantee is that the cost of the iterates is non-increasing.

The Hessian-based approach of Lefkimiatis et al. [32] and the isotropic higher-order TV approach of Hu and Jacob [28] also boil down to minimizing a function of the form (8). The Hessian-based regularization functionals are of the form $\sum_l \|\mathcal{Q}_l x\|_2$ (Frobenius-norm regularizer) and $\|\Delta x\|_1 + \sum_l \|\mathcal{Q}_l x\|_2$ (spectral-norm regularizer), where Δ is a discrete approximation to the Laplacian operator and the \mathcal{Q}_l 's are second-order differential operators ($\mathcal{Q}_l \in \mathcal{L}_{K,3}(\mathbb{R})$ for the Frobenius norm and $\mathcal{Q}_l \in \mathcal{L}_{K,2}(\mathbb{R})$ for the spectral norm). The isotropic higher-order TV regularization functional is of the form $\sum_l \|\mathcal{Q}_l x\|_q$, where the \mathcal{Q}_l 's are discrete n th-order differential operators in $\mathcal{L}_{K,n+1}(\mathbb{R})$ ($n = 2$ or 3) and $\|\cdot\|_q$ is the norm induced by a positive-definite quadratic form on \mathbb{R}^{n+1} . The resulting optimization problems are solved by majorization-minimization algorithms, but neither paper provides a convergence proof. Hu and Jacob [28] also introduced an anisotropic higher-order TV penalty to improve contour regularity while preserving edges (this penalty does not have a closed-form expression). Their experiments on image denoising show that anisotropic regularization slightly improves the SNR compared to isotropic regularization (the gain in SNR remains below 2.15%), but there is no clear evidence of edge continuation.

Other higher-order models were proposed in [10, 35, 1, 53] to reduce the staircase effect; the associated optimization algorithms are restricted to image denoising. Chan et al. [10] combined a first-order TV penalty and a second-order term that

penalizes the square of the laplacian at the locations where the gradient is small. Lysaker and Tai [35] defined the solution as a convex combination of the solutions associated with first- and second-order TV regularization. Bae et al. [1] used the Euler's elastica model [9], and Zhu and Chan [53] employed the TV of the mean curvature of the image surface. In [10, 35, 53], the optimization method is based on solving the Euler-Lagrange equations associated with the model considered, which are quite complex nonlinear PDEs. These equations are difficult to solve numerically and the proposed iterative solution schemes have very weak convergence guarantees. The approach of Bae et al. [1] is radically different: they designed an optimization algorithm based on graph cuts. Their algorithm is very efficient for image denoising, but it cannot be extended to the case where \mathcal{H} is not the identity; the reason is that the cost function is not submodular if \mathcal{H} is a blurring operator. (An alternative would be to use quadratic pseudo-Boolean optimization [48, 31], which is an extension of graph cuts that allows to deal with non-submodular cost functions. However, as discussed in [45], reconstruction techniques based on quadratic pseudo-Boolean optimization are limited to small-size blurring kernels.)

2.2. Methods using a non-quadratic data-fidelity term. The approaches reviewed above share a common limitation: the quadratic data-fidelity term $\|\mathcal{H}x - d\|_2^2$ is appropriate only for Gaussian noise and leads to solutions that are significantly altered by outliers. In particular, quadratic data-fidelity does not allow to deal with impulse noise. To overcome this drawback, a standard choice is the ℓ_1 -norm of the residual, or, more generally, a data-fidelity term of the form (3) (it is common practice to replace the ℓ_1 -norm of the residual by (3) with $\phi_k = \psi_{\text{id}}^\delta$).

Nikolova [41] proposed a convergent method to minimize cost functions of the form

$$(9) \quad U(x) = \|x - d\|_1 + \lambda \sum_l \psi(|\mathcal{R}_l x|),$$

where the \mathcal{R}_l 's are first-order difference operators and ψ is smooth at zero and convex. This approach is expressly restricted to image denoising and impulse noise; it is based on the assumptions that only a part of the pixels are corrupted by noise and that each noisy pixel is dissimilar to its nearest neighbors. Cai et al. [7] and Chan et al. [8] also considered the special case of impulse noise. They proposed a two-phase approach which extends the work of Nikolova: the noisy data pixels are first detected using an impulse-noise detector, and then the original image is estimated by minimizing

$$(10) \quad U(x) = \sum_{k \in \mathcal{J}} |(\mathcal{H}x - d)_k| + \lambda \sum_l \|\mathcal{R}_l x\|_2,$$

where the first sum is over the set \mathcal{J} of indices of the noise-free data pixels. In [7], the set $\{\mathcal{R}_l\}_l$ consists of first-order difference operators, and a smooth approximation of U is minimized using a fixed-point iteration scheme without convergence guarantee. In [8], $\{\mathcal{R}_l\}_l$ is a discrete approximation to the gradient, and U is minimized by a variant of the primal-dual method proposed in [20]. By extension, the iterative reweighted norm approach of Rodríguez and Wohlberg [47] allows to consider cost functions of the form

$$(11) \quad U(x) = \sum_{k \in \mathcal{J}} |(\mathcal{H}x - d)_k|^p + \lambda \sum_l \|\mathcal{R}_l x\|_2^q$$

with $p, q \in (0, 2)$. The associated convergence guarantees are the same as in [13]: the cost of the iterates is non-increasing, and the gradient of U goes to zero as the number of iterations grows to infinity. Li et al. [34] considered a two-phase approach to deal with Gaussian plus impulse noise: after detection of the pixels corrupted by impulse noise, the original image is estimated from the non-outlier pixels by minimizing

$$(12) \quad U(x) = \sum_{k \in \mathcal{I}} \phi_\nu(|(\mathcal{H}x - d)_k|) + \lambda \sum_l |\mathcal{R}_l x|,$$

where $\phi_\nu(t) = \nu t^2 / (\nu + t)$ and $\{\mathcal{R}_l\}_l$ is a set of tight framelet filters ($\{\mathcal{R}_l\}_l$ is the union of (i) a weighted-averaging operator, (ii) edge-detection operators, and (iii) second-order difference operators). The authors designed an iterative shrinkage algorithm in which the hyper-parameters ν and λ are determined automatically, but they did not provide any convergence guarantee. Finally, Tai et al. [49] proposed an augmented Lagrangian method for minimizing an Euler's energy functional combined with a data-fidelity term of the form $\sum_k |x_k - d_k|^p$ with $p \in [1, 2]$. Their numerical experiments demonstrate the efficiency of their algorithm, but they did not study its theoretical convergence properties.

2.3. Position of our approach. The methods reviewed above either have weak convergence guarantees [10, 35, 47, 7, 43, 1, 13, 27, 34, 49, 32, 28, 53] or are limited to very specific situations [41, 8]. Furthermore, the methods aimed at improving contour regularity and/or reducing the staircase effect are generally restricted to quadratic data-fidelity [10, 35, 1, 53, 32, 28], and the methods that employ a non-quadratic data-fidelity term are often limited to first-order models [41, 47, 7, 8]. By contrast, our approach described in the next two sections has all three of the following advantages:

- it is not limited to quadratic data-fidelity;
- it has good edge-continuation capabilities (as our experiments show);
- it offers strong convergence guarantees.

As mentioned in the introduction, the latter advantage is balanced by assumptions that rule out PFs that are either non-smooth at zero or bounded (or both), but the resulting practical limitations are minor. In addition, our deterministic relaxation algorithm can incorporate inequality constraints as well as continuation sequences at no cost. It is also easy to implement and quite generic (it can be used to minimize the cost functions considered in [41, 7, 8, 13, 27, 34, 32, 28]).

3. Implicit edge-continuation.

3.1. General description. Since we are interested in the recovery of piecewise-smooth images, we can assume that the level curves of the original continuous image (from which x^* is sampled) are piecewise-smooth. In the domain of a multiscale directional transform (such as a wavelet or a curvelet transform), this prior information is synonymous with clean and strongly oriented detail images. Consequently, edge continuation can be incorporated into the reconstruction process by means of an additional penalty term of the form

$$(13) \quad \tilde{\Psi}(\tilde{x}) = \sum_m \tilde{\psi}_m(|\mathcal{D}_m \tilde{x}|),$$

where \tilde{x} denotes the multiresolution decomposition of x , the \mathcal{D}_m 's are discrete directional derivative operators acting on the detail coefficients, and the functions $\tilde{\psi}_m$ are increasing in \mathbb{R}_+ .

Let $\mathcal{T} \in \mathcal{L}_{K, \tilde{K}}(\mathbb{R})$ be the considered multiresolution transform ($\tilde{K} \geq K$ with strict inequality if \mathcal{T} is redundant). Our estimate of x^* is defined as any global minimum of the augmented cost function $V : \mathbb{R}^K \rightarrow \mathbb{R}$ obtained by combining the contour-line smoothing term $\tilde{\Psi}$ with the standard cost function U :

$$(14) \quad V(x) = \Phi_d(x) + \lambda\Psi(x) + \tilde{\lambda}\tilde{\Psi}(\mathcal{T}x),$$

where Φ_d and Ψ are defined by (3) and (4), and where the positive parameters λ and $\tilde{\lambda}$ control the degree of smoothing in the spatial domain and in the transformed domain. The interests of this approach are that (i) it avoids the delicate use of explicit edge-processes, (ii) it can deal with edges at different resolution levels, (iii) it involves only one extra hyper-parameter (namely, $\tilde{\lambda}$), and (iv) the associated increase in optimization complexity is small if the $\tilde{\psi}_m$'s are convex.

Remark 1. If the \mathcal{D}_m 's are replaced by canonical projections, and if $\tilde{\psi}_m(t) \propto t^p$ for all m and for some fixed $p \in (0, 2)$, then $\tilde{\Psi}$ is an ℓ_p -penalty that favors sparse representations in the transformed domain. Therefore, our approach generalizes the sparsity-constraint approaches in [4, 15], which are restricted to cost functions of the form

$$(15) \quad x \in \mathbb{R}^K \mapsto \|\mathcal{H}x - d\|_2^2 + \sum_m \tilde{\lambda}_m |(\mathcal{T}x)_m|^p,$$

where the $\tilde{\lambda}_m$'s are non-negative reals. (When \mathcal{T} is a wavelet transform, ℓ_p -penalties are obtained under the assumption that the detail coefficients follow a generalized Gaussian distribution with shape parameter p [4], or when the continuous original image is assumed to belong to a Besov space [15].)

Remark 2. The issue of choosing appropriate values for λ and $\tilde{\lambda}$ is beyond the scope of this paper. Good starting points are the L-hypersurface method [5], the Monte Carlo SURE approach [44], the use of the no-reference measure proposed in [54], and the zero-crossing method developed in [30]. (The L-hypersurface method requires the PFs to be convex, the approaches in [44] and [54] are restricted to image denoising, and the method in [30] must be calibrated as a function of the regularization model.) Another avenue is to determine the parameters of the cost function automatically by updating their values adaptively during the course of the optimization process, as in [34], but this considerably complicates the convergence analysis.

In the following section, we refine the description of the contour-line smoothing term $\tilde{\Psi}$ in the case of a non-redundant wavelet transform. We focus on this situation for computational efficiency reasons and because the construction is easier when directional selectivity is limited. However, the approach can be applied to the case of a redundant transform with high directional selectivity (for instance, the curvelet transform [36]) at the expense of additional design efforts and increased computational burden.

3.2. Wavelet-domain edge-continuation. Let $J \in \mathbb{N}^*$. The J -level decomposition of an image $x \in \mathbb{R}^K$ in a wavelet basis is a set of subimages

$$\{ \mathcal{A}_0 x, \mathcal{A}_j^h x, \mathcal{A}_j^v x, \mathcal{A}_j^d x : j \in \llbracket 0, J-1 \rrbracket \},$$

where the index j represents the scale, $\mathcal{A}_0 x$ is the approximation of x at resolution 2^{-J} (the resolution of x is equal to one and its scale is zero), and the detail images $\mathcal{A}_j^h x$, $\mathcal{A}_j^v x$, and $\mathcal{A}_j^d x$ convey the difference of information between $\mathcal{A}_j x$ (the approximation of x at resolution 2^{j-J}) and the finer approximation $\mathcal{A}_{j+1} x$. The linear map \mathcal{A}_j^h gives the horizontal high frequencies (that is, the vertical edges), \mathcal{A}_j^v gives the vertical high frequencies (that is, the horizontal edges), and \mathcal{A}_j^d gives the high frequencies in both horizontal and vertical directions. The wavelet transform \mathcal{T} can be viewed as the concatenation of the component functions of the maps $\mathcal{A}_0 \in \mathcal{L}_{K,4^{-J}K}(\mathbb{R})$ and $\mathcal{A}_j^h, \mathcal{A}_j^v, \mathcal{A}_j^d \in \mathcal{L}_{K,4^{j-J}K}(\mathbb{R})$ for $j \in \llbracket 0, J-1 \rrbracket$.

Piecewise-smooth level curves in the original continuous image translate to the fact that the horizontal and vertical detail images of x^* are respectively vertically and horizontally oriented. In other words, the vertical derivatives in $\mathcal{A}_j^h x^*$ and the horizontal derivatives in $\mathcal{A}_j^v x^*$ have small amplitude. This suggests to define the contour-line smoothing term as follows:

$$(16) \quad \tilde{\Psi}(\mathcal{T}x) = \sum_{j \in \llbracket 0, J-1 \rrbracket} \left(\sum_{k \in \llbracket 1, K_j^v \rrbracket} \tilde{\psi}(\alpha_j |D_{j,k}^v \mathcal{A}_j^h x|) + \sum_{k \in \llbracket 1, K_j^h \rrbracket} \tilde{\psi}(\alpha_j |D_{j,k}^h \mathcal{A}_j^v x|) \right),$$

where the α_j 's are positive reals, and where the D_j^v 's and the D_j^h 's are vertical and horizontal discrete derivative operators, respectively. The outer summation is over scales, and the inner summations are over the pixels of the detail images where the derivatives are evaluated. A natural choice for the PF $\tilde{\psi}$ is the identity on \mathbb{R}_+ , for we want to preserve both sharp transitions and smooth variations along the preferential directions of the detail images. In practice, we advocate using a strictly convex function such as (5), which facilitates the optimization process without compromising the quality of the reconstructions. The weights α_j allow to strengthen the edge-continuation effect at some particular resolution level(s). If there is no conflict with possible application-specific constraints, we choose the α_j 's so as to compensate for wavelet-coefficient decay, as discussed in Section 3.3.

From a filtering perspective, the action of $\tilde{\Psi}$ is to discard the solutions that convey significant information in the frequency regions schematized in Fig. 1(a). These regions are exemplified in Figs. 1(b)–(d) for the setting used in our experiments (namely, biorthogonal spline wavelets with two vanishing moments [14], and first-order difference derivative operators). It appears that the proposed wavelet-domain smoothing term not only favors the formation of horizontal and vertical edges, but also preserves diagonal boundaries (which is why we do not add a penalty on the diagonal subbands $\mathcal{A}_j^d x$, $j \in \llbracket 0, J-1 \rrbracket$).

Remark 3. Although the parallel is tempting, $\tilde{\Psi}$ is not comparable to a second- or higher-order smoothing term in the spatial domain. To fix ideas, consider the standard second-order model introduced in [22]:

$$(17) \quad \Psi_2 = \Psi^h + \Psi^v + \Psi^d \quad \text{with} \quad \Psi^x(x) = \sum_{l \in \llbracket 1, L^x \rrbracket} \psi(|\mathcal{R}_l^x x|),$$

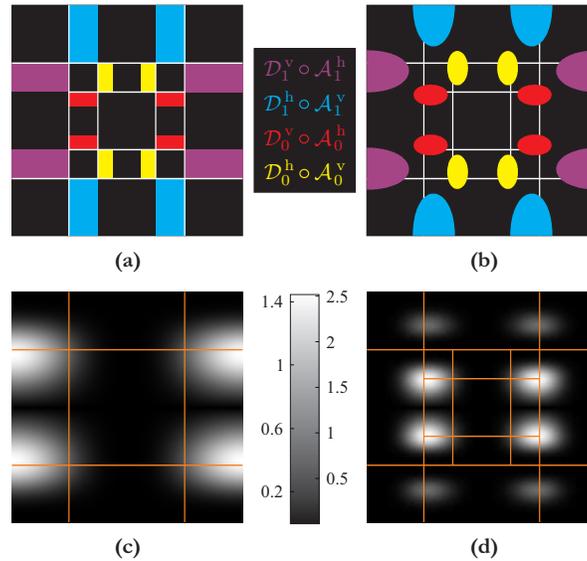


FIGURE 1. Frequency regions penalized by the wavelet-domain smoothing term $\tilde{\Psi}$ defined in (16) for $J = 2$: (a) schematization of the passbands of the involved filtering operators; (b)–(d) 3dB passbands and frequency responses for biorthogonal spline wavelets and first-order differences (Figs. 1(c) and 1(d) only show the frequency responses of $\mathcal{D}_1^v \mathcal{A}_1^h$ and $\mathcal{D}_0^v \mathcal{A}_0^h$).

where x stands for any of the letters $h, v,$ and $d,$ and where the second-derivative operators $\mathcal{R}^h, \mathcal{R}^v,$ and \mathcal{R}^d are the convolutions with the masks

$$(18) \quad M^h = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}, \quad M^v = (M^h)^T, \quad \text{and} \quad M^d = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Fig. 2 displays the frequency responses of \mathcal{R}^h and \mathcal{R}^d (the response of \mathcal{R}^v is the 90° rotation of that of \mathcal{R}^h)—it shows that Ψ_2 penalizes the high-frequency components regardless of their orientation in the spatial domain. Indeed, \mathcal{R}^h and \mathcal{R}^v filter out the horizontal and the vertical high frequencies, including the diagonal subbands which are further attenuated by \mathcal{R}^d . Therefore, the combined effect of $\Psi^h, \Psi^v,$ and Ψ^d (or of Ψ^h and Ψ^v only) is that of a low-pass filter with maximum attenuation in the diagonal subbands, which reveals that second-order models are not even appropriate for edge-preservation. The same conclusion holds for third- and higher-order models, as the strength of spatial smoothing increases with the order of the model.

3.3. Correcting wavelet-coefficient decay. We propose to set the coefficients α_j in (16) so as to compensate for the decay of the magnitudes of the detail coefficients across scales. An elegant way to do so is to lie at the intersection of two standard image models: the independent generalized Gaussian (IGG) model [37, 40, 4], and the Besov space (BS) model [18, 15].

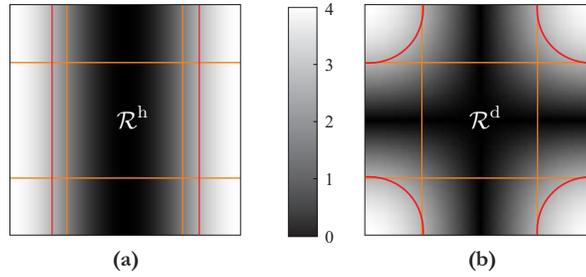


FIGURE 2. Frequency responses of the second-order derivative operators \mathcal{R}^h and \mathcal{R}^d defined by the masks M^h and M^d given in (18) (the red contour lines delimit the 3 dB passbands).

The IGG model is a statistical model in which it is assumed that (i) the wavelet coefficients are independent, and (ii) the detail coefficients at scale j follow a zero-mean, generalized Gaussian distribution π_j with shape parameter $p_j > 0$ and variance σ_j^2 :

$$(19) \quad \pi_j(t) \propto \exp\left(- (c_j|t|/\sigma_j)^{p_j}\right),$$

where $c_j = (\Gamma(3/p_j)/\Gamma(1/p_j))^{1/2}$ (Γ is the standard Gamma function). From the standpoint of this model, wavelet decay correction is achieved by multiplying the \mathcal{D}_j^y 's and the \mathcal{D}_j^h 's by σ_0/σ_j . To obtain a guideline for setting the ratios σ_0/σ_j , we take inspiration from [15] by assuming that the original discrete image x^* is sampled from a continuous image $f^* : D \rightarrow \mathbb{R}$ (D is a rectangular domain of \mathbb{R}^2) that belongs to a Besov space $B_p^s(L_p(D))$ with $p \geq 1$ and $s \geq 2/p - 1$. The precise definitions of $B_p^s(L_p(D))$ and of its standard norm $\|\cdot\|_{B_p^s(L_p(D))}$ are given in [18]. All we need to know here is that the BS model is a deterministic characterization of image smoothness: $f^* \in B_p^s(L_p(D))$ essentially means that f^* has “ s derivatives” in $L_p(D)$.

The relationship between the IGG and BS models arises when

$$(20) \quad \forall j \geq 0, \quad \begin{cases} p_j = p, \\ \sigma_j = 2^{-j(s+1-2/p)}\sigma_0. \end{cases}$$

(In other words, we further assume that the IGG shape parameter is constant and that the wavelet coefficients decay exponentially across scales.) In this case, the standard Besov norm on $B_p^s(L_p(D))$ is equivalent to the negative logarithm of the normalized IGG likelihood raised to the power of $1/p$ [12]; that is, letting $\{w_{j,k}^x(f) : k \in \mathbb{Z}\}$ be the set of detail coefficients of $f : D \rightarrow \mathbb{R}$ at scale j and at orientation $x = h, v, \text{ or } d$, we have

$$(21) \quad \begin{aligned} \|f\|_{B_p^s(L_p(D))} &\sim \left(-\ln\left(\prod_{j \geq 0, k, x} \frac{\pi_j(w_{j,k}^x(f))}{\pi_j(0)}\right)\right)^{1/p} \\ &\propto \left(\sum_{j \geq 0, k, x} 2^{jp(s+1-2/p)} |w_{j,k}^x(f)|^p\right)^{1/p}. \end{aligned}$$

Hence, if (20) is satisfied, the more likely an image according to the IGG model, the smaller its Besov norm (and vice versa). This connection between the IGG and BS models suggests to set $\alpha_j = \sigma_0/\sigma_j = 2^{j(s+1-2/p)}$ for some values of s and p that reflect some knowledge about the original continuous image f^* in addition to (20). In our experiments (see Section 5), we set $p = 1$ and $s = 2$ (that is, $\alpha_j = 2^j$), which amounts to assume that f^* has second-order smoothness in $L_1(D)$ and that it conforms to the independent Laplacian model with standard deviation $2^{-j}\sigma_0$ at scale j .

4. Deterministic relaxation. In this section, we focus on the minimization of the augmented cost function V given in (14) (the proposed wavelet-domain smoothing term defined in (16) is an instance of (13)). For convenience, we write V in the more compact form

$$(22) \quad V(x) = \Phi_d(x) + \sum_{l \in \llbracket 1, L' \rrbracket} \theta_l(\|\mathcal{Q}_l x\|_2),$$

where $L' = L + M$ and

$$(23) \quad (\theta_l, \mathcal{Q}_l) = \begin{cases} (\lambda \psi_l, \mathcal{R}_l) & \text{if } l \in \llbracket 1, L \rrbracket, \\ (\tilde{\lambda} \tilde{\psi}_{l-L}, \mathcal{D}_{l-L} \mathcal{T}) & \text{if } l \in \llbracket L + 1, L' \rrbracket. \end{cases}$$

We assume that the data penalty functions $\phi_1, \dots, \phi_{K'}$ and the PFs $\theta_1, \dots, \theta_{L'}$ satisfy the following conditions:

$$\text{C1. } \underbrace{(\forall k \in \llbracket 1, K' \rrbracket, \phi_k \in \mathfrak{S})}_{\text{C1.1}} \text{ OR } \underbrace{(\forall k \in \llbracket 1, K' \rrbracket, \phi_k(t) = t^2)}_{\text{C1.2}}$$

$$\text{C2. } \forall l \in \llbracket 1, L' \rrbracket, \theta_l \in \mathfrak{S}$$

where \mathfrak{S} is the set of functions $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that

- (a) $\varphi(0) = 0$, φ is increasing, and φ is twice differentiable;
- (b) $\lim_{t \rightarrow +\infty} \varphi(t) = +\infty$;
- (c) φ is four times differentiable at zero and $\varphi'''(0) = 0$;
- (d) the function

$$(24) \quad \varphi^\dagger : t \in \mathbb{R}_+ \mapsto \begin{cases} \varphi'(t)/t & \text{if } t > 0, \\ \varphi''(0) & \text{if } t = 0, \end{cases}$$

is strictly decreasing, and $\lim_{t \rightarrow +\infty} \varphi^\dagger(t) = 0$.

Conditions of type (a) are standard in regularized reconstruction; conditions (b) ensure that V is coercive (that is, $\lim_{\|x\| \rightarrow \infty} V(x) = +\infty$) and hence has a global minimum; and conditions (c) are technical requirements for convergence proofs. Conditions of type (d) have different implications depending on whether they concern the data penalty functions or the PFs: their application to the ϕ_k 's allows to handle the case of a non-Gaussian noise, and their application to the θ_l 's guarantees edge-preservation properties.

Remark 4. C1.1 and C1.2 are mutually exclusive (because $\phi_k^\dagger = 2$ if $\phi_k(t) = t^2$). The choice C1.2 allows to consider the important special case when the data-fidelity term is an ℓ_2 -norm weighted by a symmetric positive-definite matrix (call it \mathcal{S}), that is, when Φ_d is of the form

$$(25) \quad \Phi_d(x) = (\mathcal{H}x - d)^T \mathcal{S}(\mathcal{H}x - d) = \|\mathcal{S}^{1/2}(\mathcal{H}x - d)\|_2^2.$$

This situation arises in the Bayesian setting when there is no impulse noise and when the noise vector η has a multivariate normal distribution with covariance matrix \mathcal{S}^{-1} .

Our deterministic relaxation algorithm for minimizing V is described in Section 4.1. Its convergence properties are presented in Section 4.2, and we discuss two possible refinements in Section 4.3 (namely, the enforcement of inequality constraints and the introduction of continuation sequences).

4.1. Construction of the algorithm. Let $e_\phi = (e_{\phi,1}, \dots, e_{\phi,K'}) : \mathbb{R}^K \rightarrow (0, +\infty)^{K'}$ and $e_\theta = (e_{\theta,1}, \dots, e_{\theta,L'}) : \mathbb{R}^K \rightarrow (0, +\infty)^{L'}$ be the vector-valued functions with components

$$(26) \quad e_{\phi,k}(x) = \phi_k^\dagger(|(\mathcal{H}x - d)_k|) \quad \text{and} \quad e_{\theta,l}(x) = \theta_l^\dagger(\|\mathcal{Q}_l x\|_2).$$

For any $x \in \mathbb{R}^K$, we define the diagonal matrices

$$(27) \quad \mathcal{E}_\phi(x) = \text{diag}(e_{\phi,k}(x) : k \in \llbracket 1, K' \rrbracket)$$

$$(28) \quad \text{and} \quad \mathcal{E}_\theta(x) = \text{diag}(e_{\theta,l}(x)\mathcal{I}_{\rho_l} : l \in \llbracket 1, L' \rrbracket),$$

where \mathcal{I}_r is the identity matrix of order r and ρ_l is the number of rows of \mathcal{Q}_l . The first-order necessary condition for $x \in \mathbb{R}^K$ to be an extremum of V (that is, $\nabla V(x) = 0$) writes

$$(29) \quad \underbrace{(\mathcal{H}^T \mathcal{E}_\phi(x) \mathcal{H} + \mathcal{Q}^T \mathcal{E}_\theta(x) \mathcal{Q})}_{=: \mathcal{M}(x)} x = \mathcal{H}^T \mathcal{E}_\phi(x) d,$$

where $\mathcal{Q} = [\mathcal{Q}_1^T, \dots, \mathcal{Q}_{L'}^T]^T$ is the vertical concatenation of the \mathcal{Q}_l 's. For any $x \in \mathbb{R}^K$, the symmetric matrix $\mathcal{M}(x)$ is positive semidefinite, and it is positive-definite if and only if

$$\mathbf{C3.} \quad \ker(\mathcal{H}) \cap \ker(\mathcal{Q}) = \{0\}.$$

Assuming this condition holds, (29) suggests the following iterative relaxation algorithm:

$$(30) \quad \mathbb{A}_V \equiv \begin{cases} x^{(0)} \in \mathbb{R}^K, \\ \forall n \in \mathbb{N}, \quad x^{(n+1)} = (\mathcal{M}(x^{(n)}))^{-1} \mathcal{H}^T \mathcal{E}_\phi(x^{(n)}) d. \end{cases}$$

This fixed-point iteration scheme can be equivalently written as

$$(31) \quad x^{(n+1)} = \arg \inf_{x \in \mathbb{R}^K} V_0(x, e_\phi(x^{(n)}), e_\theta(x^{(n)})),$$

where $V_0 : \mathbb{R}^K \times \mathbb{R}^{K'} \times \mathbb{R}^{L'} \rightarrow \mathbb{R}$ is defined by

$$(32) \quad V_0(x, \delta, \varepsilon) = \sum_{k \in \llbracket 1, K' \rrbracket} \delta_k (\mathcal{H}x - d)_k^2 + \sum_{l \in \llbracket 1, L' \rrbracket} \varepsilon_l \|\mathcal{Q}_l x\|_2^2.$$

Condition C3 guarantees that the quadratic function $x \in \mathbb{R}^K \mapsto V_0(x, \delta, \varepsilon)$ is positive-definite for any $(\delta, \varepsilon) \in (0, +\infty)^{K'} \times (0, +\infty)^{L'}$, and thus that the half-quadratic algorithm \mathbb{A}_V can be efficiently implemented using a preconditioned conjugate gradient method. Condition C3 trivially holds when $\text{rank}(\mathcal{H}) = K$, as in image denoising problems. It is also usually satisfied for reconstruction problems where $\text{rank}(\mathcal{H}) < K$, such as image deblurring and limited-angle tomography. Indeed, piecewise-smooth reconstruction involves gradients or directional derivatives

in the spatial domain, which implies that $\ker(\mathcal{Q}) = \text{span}\{\mathbf{1}_K\}$ ($\mathbf{1}_K$ denotes the vector in \mathbb{R}^K whose entries are all equal to one), and we generally have $\mathcal{H}\mathbf{1}_K \neq 0$.

4.2. Convergence properties. The convergence properties of \mathbb{A}_V (30) are summarized in Theorem 4.1 below, whose proof is given in Appendix A. We denote by \mathfrak{C} the set of critical points of V , that is,

$$(33) \quad \mathfrak{C} = \{x \in \mathbb{R}^K \mid \nabla V(x) = 0\},$$

and we use the following terminology: a *continuum* of \mathbb{R}^K is a connected compact subset of \mathbb{R}^K ; we call $x \in \mathfrak{C}$ *isolated* if

$$(34) \quad \exists a > 0, \forall y \in \mathfrak{C} \setminus \{x\}, \|y - x\| \geq a$$

($\|\cdot\|$ is any norm on \mathbb{R}^K); and \mathfrak{C} is said to be *discrete* if all its points are isolated.

Theorem 4.1. *Assume that conditions C1–C3 are satisfied, and let $(x^{(n)})_n$ be any sequence generated by \mathbb{A}_V .*

- (I) $(x^{(n)})_n$ has convergent subsequences, and all its accumulation points are in \mathfrak{C} .
- (II) There exists $x^* \in \mathfrak{C}$ such that $\lim_{n \rightarrow \infty} V(x^{(n)}) = V(x^*)$, and, for any $n \in \mathbb{N}$, $V(x^{(n+1)}) \leq V(x^{(n)})$ with equality if and only if $x^{(n)} \in \mathfrak{C}$.
- (III) Either $(x^{(n)})_n$ converges, or its accumulation points form a continuum.
- (IV) Let x^* be an isolated critical point of V . If x^* is a local minimum of V , then there exists an open neighborhood \mathfrak{N} of x^* such that $\lim_{n \rightarrow \infty} x^{(n)} = x^*$ whenever $x^{(0)} \in \mathfrak{N}$.
- (V) $\lim_{n \rightarrow \infty} \inf_{y \in \mathfrak{C}} \|x^{(n)} - y\| = 0$.
- (VI) If \mathfrak{C} is discrete, then there exists $x^* \in \mathfrak{C}$ such that $\lim_{n \rightarrow \infty} x^{(n)} = x^*$.
- (VII) If V is strictly convex, then $(x^{(n)})_n$ converges to the global minimum of V .

Theorem 4.1 characterizes the behavior of \mathbb{A}_V in three mutually exclusive situations which cover all possibilities: (i) when V is strictly convex, (ii) when V is not strictly convex and \mathfrak{C} is discrete, and (iii) when \mathfrak{C} is not discrete. The corresponding conclusions are the following:

- (i) The cost function V is strictly convex if C3 holds and if the data penalty functions $\phi_1, \dots, \phi_{K'}$ and the PFs $\theta_1, \dots, \theta_{L'}$ are strictly convex. Therefore, from (VII), convergence to the global minimum always occurs when the ϕ_k 's and the θ_l 's are strictly convex.
- (ii) If V is not strictly convex and \mathfrak{C} is discrete, then we have from (VI) that $(x^{(n)})_n$ converges to a critical point $x^* \in \mathfrak{C}$. We cannot rigorously exclude the possibility that x^* is a maximum or a saddle point; however, this is very unlikely because, while any isolated critical point that is a minimum is an attractor (by (IV)), maxima and saddle points are unstable. Indeed, from (II), small random perturbations away from an isolated critical point that is not a minimum will eventually move the iterates away from it. Therefore, from an algorithmic standpoint, convergence to a maximum or a saddle point can be avoided by adding a small random vector to the current solution whenever $\|x^{(n+1)} - x^{(n)}\|/\|x^{(n)}\|$ is below a certain threshold. In practice, the round-off errors in floating-point arithmetic do the job, and hence any numerical implementation will lead to a minimum.
- (iii) If \mathfrak{C} is not discrete, then, from (I) and (III), either $(x^{(n)})_n$ converges, or \mathfrak{C} contains a non-empty continuum. The latter case may be considered as a failure in the design of the regularization term(s), but the algorithm also

behaves well in this situation: by (II) and (v), $(V(x^{(n)}))_n$ converges, and $(x^{(n)})_n$ gets arbitrarily close to \mathfrak{C} .

4.3. Refinements.

4.3.1. *Inequality constraints.* The proposed optimization scheme can be easily extended to enforce inequality constraints on the pixel values, that is, constraints of the form $\mathcal{A}x \geq b$ with $\mathcal{A} \in \mathcal{L}_{K,M'}(\mathbb{R})$ and $b \in \mathbb{R}^{M'}$. Indeed, it suffices to consider the cost function

$$(35) \quad W(x) = V(x) + \mu \|(b - \mathcal{A}x)^+\|_2^2,$$

where μ is a positive real constant and y^+ denotes the positive part of y (that is, $y^+ = (y_1^+, \dots, y_{M'}^+)$ with $y_m^+ = \max\{y_m, 0\}$).

Assume that conditions C1–C3 hold, and let $W_0 : \mathbb{R}^K \times \mathbb{R}^{K'} \times \mathbb{R}^{L'} \rightarrow \mathbb{R}$ be the function defined by

$$(36) \quad W_0(x, \delta, \varepsilon) = V_0(x, \delta, \varepsilon) + 2\mu \|(b - \mathcal{A}x)^+\|_2^2,$$

where V_0 is given in (32). The quadratic function $x \in \mathbb{R}^K \mapsto V_0(x, \delta, \varepsilon)$ is strictly convex and coercive for any $(\delta, \varepsilon) \in (0, +\infty)^{K'} \times (0, +\infty)^{L'}$, and the penalty function $x \in \mathbb{R}^K \mapsto \|(b - \mathcal{A}x)^+\|_2^2$ is continuously differentiable and convex. Therefore, for any positive δ and ε , the piecewise-quadratic function $W_0(\cdot, \delta, \varepsilon) : x \in \mathbb{R}^K \mapsto W_0(x, \delta, \varepsilon)$ is continuously differentiable, strictly convex, and coercive, and thus it has a unique global minimum which can be easily found (using, for instance, a conjugate gradient method). This leads to the following extension of \mathbb{A}_V :

$$(37) \quad \mathbb{A}_W \equiv \begin{cases} x^{(0)} \in \mathbb{R}^K, \\ \forall n \in \mathbb{N}, x^{(n+1)} = \arg \inf_{x \in \mathbb{R}^K} W_0(x, e_\phi(x^{(n)}), e_\theta(x^{(n)})), \end{cases}$$

where e_ϕ and e_θ are given in (26). As is made clear in Appendix A, all the convergence results discussed in Section 4.2 generalize to W along with \mathbb{A}_W .

4.3.2. *Continuation.* The algorithm \mathbb{A}_V is very efficient when the data penalty functions $\phi_1, \dots, \phi_{K'}$ and the PFs $\theta_1, \dots, \theta_{L'}$ are strictly convex and when their derivatives do not increase too rapidly at the origin. On the other hand, \mathbb{A}_V may be trapped in a poor local minimum when the ϕ_k 's or the θ_l 's are not convex, and its convergence is slow when the ϕ_k 's or the θ_l 's closely approximate a function that is not smooth at zero. In these latter two situations, the convergence can be improved by replacing some or all of the ϕ_k 's and of the θ_l 's with the elements of some carefully designed sequences $(\phi_k^{(n)})_{n \in [1, N]}$ and $(\theta_l^{(n)})_{n \in [1, N]}$ at the beginning of the optimization process. The resulting modified algorithm is the same as \mathbb{A}_V , except for the first N iterations where \mathcal{E}_ϕ and \mathcal{E}_θ are respectively replaced by $\mathcal{E}_{\phi^{(n)}}$ and $\mathcal{E}_{\theta^{(n)}}$. This continuation technique can be thought of as providing a good initialization for \mathbb{A}_V , and the convergence results in Theorem 4.1 still hold.

To actually improve the performance of \mathbb{A}_V , the sequences $(\phi_k^{(n)})_n$ and $(\theta_l^{(n)})_n$ should satisfy conditions C4–C6 below. We let $(\varphi^{(n)})_n$ be one of these sequences, $\varphi = \phi_k$ or θ_l be the associated target function, and $\kappa(\varphi^{(n)})$ be the *maximum concavity* of $\varphi^{(n)}$, that is,

$$(38) \quad \kappa(\varphi^{(n)}) = \sup_{t \geq 0} \max\{-(\varphi^{(n)})''(t), 0\}.$$

C4. $\varphi^{(n)}$ should be as close as possible as φ for all n .

C5. If φ approximates a function that is not smooth at zero, then the sequence

$((\varphi^{(n)})''(0))_n$ should be gradually increasing to $\varphi''(0)$.

C6. If φ is not convex, then the sequence $(\kappa(\varphi^{(n)}))_n$ should be gradually increasing to $\kappa(\varphi)$, so that the non-convexity (and hence the difficulty) of the optimization problem increases progressively with the number of iterations.

For instance, if the target is the convex PF ψ_{id}^δ (see (5)), a natural choice is $\varphi^{(n)} = \psi_{\text{id}}^{\delta_n}$ with δ_n decreasing with n . This sequence speeds up the solution search by gradually reducing the range over which $\varphi^{(n)}$ acts as a quadratic function. If the target is the non-convex PF ψ_{HL}^δ (see (7)), we suggest to use $\varphi^{(n)} = (1 - a_n)\psi_{\text{id}}^\delta + a_n\psi_{\text{HL}}^\delta$ with a_n increasing with n from 0 to 1. In this way, the maximum concavity increases progressively from 0 to $\kappa(\psi_{\text{HL}}^\delta) = 1/(4\delta^2)$, and the algorithm can reach deeper minima. These two example sequences are used in our experiments (see Remark 5).

5. Experiments. We consider the reconstruction of the 128×128 images shown in Figs. 3(a) and 4(a) from the data displayed in Figs. 3(b) and 4(b), respectively. The data associated with the ‘‘office’’ image were generated by first blurring with a 7×7 uniform mask and then adding white Gaussian noise at 20 dB SNR. (The standard deviation σ of the noise is defined via the decibel level of the signal-to-noise ratio: $\text{SNR}_{\text{dB}} = 20 \log_{10}(\sigma^*/\sigma)$, where σ^* is the standard deviation of the exact data $\mathcal{H}(x^*)$.) The data associated with the ‘‘peppers’’ image were obtained by first blurring with a 7×7 rotationally symmetric Gaussian mask with a standard deviation of 2 pixels, then adding white Gaussian noise at 20 dB SNR, and finally adding random-valued impulse noise with a 30% corruption rate. More precisely, letting $\mathfrak{d} = \mathcal{H}x^* + \epsilon_\eta$ be the blurred image corrupted by Gaussian noise, the data $d = \chi(\mathfrak{d})$ is defined by

$$(39) \quad d_k = \begin{cases} \mathfrak{d}_k & \text{with probability 0.7,} \\ \tau_k & \text{with probability 0.3,} \end{cases}$$

where τ_k is a sample from the uniform density on $[\min_k \mathfrak{d}_k, \max_k \mathfrak{d}_k]$.

Our estimates of the original images are obtained by minimizing the cost function V given in (14), where the wavelet-domain smoothing term $\tilde{\Psi}$ is defined by (16), using the algorithm Δ_V (30) with a maximum number of 40 iterations. (We do not add constraints on the pixel value range to better evaluate the improvement brought by wavelet-domain smoothing; see Section 5.1.4.) Since the noise characteristics are stationary, the data penalty functions ϕ_k are all set equal to a same function ϕ . Likewise, since we do not use any information about possible edge-locations, the PFs ψ_l in the spatial-domain smoothing term Ψ (4) are all set equal to a same function ψ . Each operator \mathcal{R}_l in Ψ computes a discrete approximation to the gradient at pixel l ; that is, if $(l_1, l_2) \in \llbracket 1, L_1 \rrbracket \times \llbracket 1, L_2 \rrbracket$ ($L_1 = L_2 = 128$) and $x(l_1, l_2)$ respectively denote the coordinates and the value of pixel l in x , then $\mathcal{R}_l(x) \in \mathbb{R}^2$ is defined by

$$(40) \quad \mathcal{R}_l(x) = (x(l_1, l_2) - x(l_1 - 1, l_2), x(l_1, l_2) - x(l_1, l_2 - 1))$$

for any $(l_1, l_2) \in \llbracket 2, L_1 \rrbracket \times \llbracket 2, L_2 \rrbracket$, and

$$(41) \quad \mathcal{R}_l(x) = \begin{cases} (0, x(l_1, l_2) - x(l_1, l_2 - 1)) & \text{if } l_1 = 1 \text{ and } l_2 \in \llbracket 2, L_2 \rrbracket, \\ (x(l_1, l_2) - x(l_1 - 1, l_2), 0) & \text{if } l_1 \in \llbracket 2, L_1 \rrbracket \text{ and } l_2 = 1, \\ (0, 0) & \text{if } (l_1, l_2) = (1, 1). \end{cases}$$

We consider the following choices for ϕ and ψ :

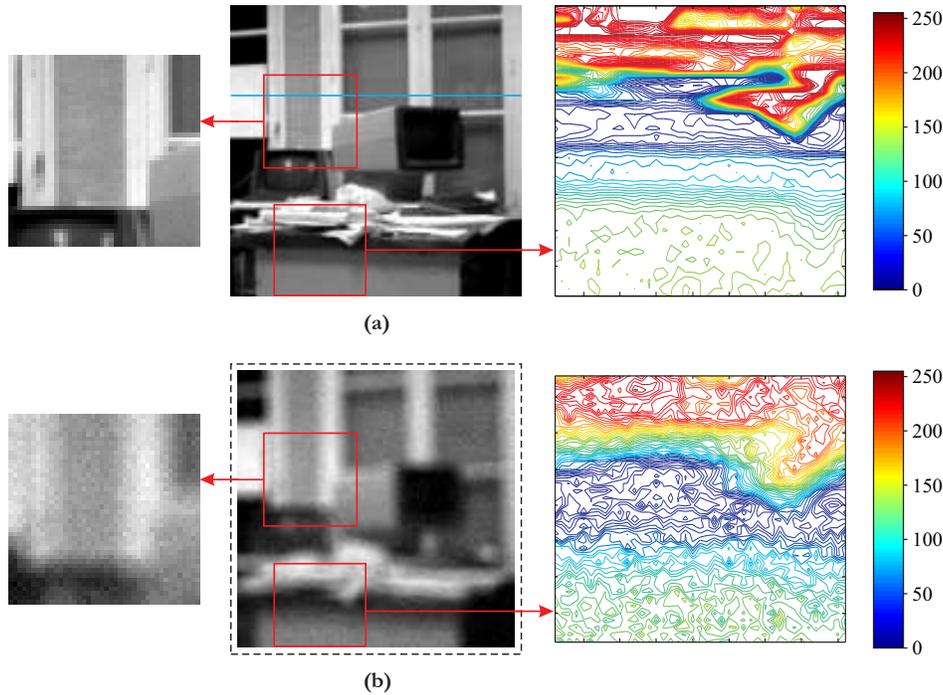


FIGURE 3. “Office” image: (a) original (the cyan horizontal line indicates the location of the profiles shown in Fig. 11); (b) degraded observation (7×7 uniform blur + 20 dB white Gaussian noise).

- $\phi(t) = t^2$, which is appropriate when the noise is additive and consists of i.i.d. normal random variables (the data-fidelity term $\Phi_d(\cdot)$ is the ℓ_2 -norm of the residual $\mathcal{H}(\cdot) - d$).
- $\phi = \psi_{\text{id}}^{0,1}$, as defined in (5), which is appropriate for dealing with impulse noise ($\Phi_d(\cdot)$ approximates the ℓ_1 -norm of $\mathcal{H}(\cdot) - d$).
- $\psi = \psi_{\text{id}}^{0,1}$, referred to as the *convex case* ($\Psi(x)$ is a smooth, convex approximation to the discrete TV of x).
- $\psi = \psi_{\text{HL}}^{10}$, as defined in (7), referred to as the *non-convex case* (Ψ acts as a quadratic regularizer for gradient magnitudes up to 3–4 and as an edge detector for gradient magnitudes greater than 10).

Remark 5. We use the continuation sequences proposed in Section 4.3.2 during the first 30 iterations of \mathbb{A}_V : $\psi_{\text{id}}^{0,1}$ is replaced by $\psi_{\text{id}}^{\delta_n}$ with $(\delta_n)_{n \in \llbracket 1, 30 \rrbracket}$ decreasing linearly from 10 to 0.1, and ψ_{HL}^{10} is replaced by $(1 - a_n)\psi_{\text{id}}^{10} + a_n\psi_{\text{HL}}^{10}$ with $(a_n)_{n \in \llbracket 1, 30 \rrbracket}$ increasing linearly from 0 to 1.

For the wavelet-domain smoothing term, we use biorthogonal spline wavelets with two vanishing moments [14] and two resolution levels (we set $\alpha_0 = 1$ and $\alpha_1 = 2$ in accordance with the discussion in Section 3.3). We chose these particular wavelets because they combine symmetry, regularity, and compact support [50]. The \mathcal{D}_j^y 's compute differences between vertically adjacent coefficients in the horizontal detail images, and the \mathcal{D}_j^h 's compute differences between horizontally adjacent coefficients in the vertical detail images. In other words, \mathcal{D}_j^y convolves $\mathcal{A}_j^h(x)$ with the mask

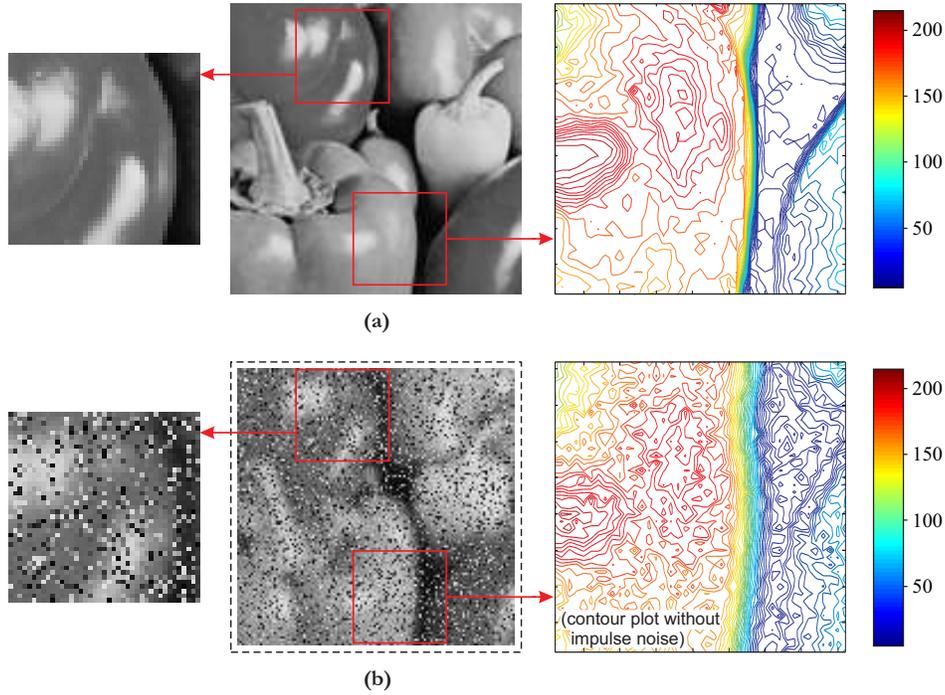


FIGURE 4. “Peppers” image: (a) original; (b) degraded observation (7×7 Gaussian blur + 20 dB white Gaussian noise + 30% random-valued impulse noise).

$[-1 \ 1]^T$, and \mathcal{D}_j^h convolves $\mathcal{A}_j^y(x)$ with the mask $[-1 \ 1]$ (there are of course other and better discrete approximations of a derivative [51, 21], but the associated extra computational load does not worth the effort). Finally, the PF $\tilde{\psi}$ is set equal to $\psi_{\text{id}}^{0,1}$, so that the inner summations in $\tilde{\Psi}$ approximate the vertical TVs of the horizontal detail images and the horizontal TVs of the vertical detail images.

Our results are described in the next two sections, where our estimates are compared with those obtained using the following state-of-the-art methods:

- iterative framelet-based approximation/sparsity (IFAS) reconstruction [34],
- Hessian Frobenius-norm (HFN) regularization [32],
- isotropic second-order TV (ITV2) regularization [28].

(We also performed experiments using the Hessian spectral-norm regularizer proposed in [32], but we do not present the corresponding results because there are slightly less good than those achieved with the HFN regularizer.) The metrics used to assess reconstruction quality are the mean-square error (MSE) and the *improvement in SNR* (ISNR) [2]; the ISNR associated with a computed solution \hat{x} is defined by

$$(42) \quad \text{ISNR}_{\text{dB}} = 20 \log_{10} \left(\frac{\|x^*|_{S'} - d\|_2}{\|x^*|_{S'} - \hat{x}|_{S'}\|_2} \right),$$

where $x|_{S'}$ denotes the restriction of x to the lattice S' supporting d (in image deblurring, the solution space \mathbb{R}^K and the data space $\mathbb{R}^{K'}$ are supported by two rectangular lattices S and S' such that $S' \subset S$).

5.1. **“Office” image.** Since the blurred observation of the “office” image is corrupted by additive white Gaussian noise only, we restrict ourselves here to quadratic data-fidelity (that is, $\phi(t) = t^2$).

5.1.1. *Convex case.* Fig. 5(a) shows the best reconstruction obtained with the PF $\psi = \psi_{\text{id}}^{0,1}$ and without wavelet-domain smoothing. The corresponding value of the spatial-domain smoothing parameter λ is 1.0, the MSE is 208.0, and the ISNR is 5.70dB. The edges are noisy (not to say visually unpleasant), and the staircase effect is clearly visible at the bottom of the image. By comparison, Fig. 5(b) displays the best solution achieved by adding the wavelet-domain smoothing term and keeping $\lambda = 1.0$ (this solution is obtained for $\tilde{\lambda} = 0.8$). There are noticeable improvements: the image contours are smoother without penalizing object boundaries, and the “patchy” appearance due to the staircase effect is softened. Quantitatively, the MSE is 180.4 and the ISNR is 6.32dB. Fig. 6 displays the ISNR as a function of the wavelet-domain smoothing parameter $\tilde{\lambda}$. Too large values of $\tilde{\lambda}$ predictably produce over-smoothed solutions, but the values of $\tilde{\lambda}$ that improve reconstruction quality cover a large interval compared to λ .

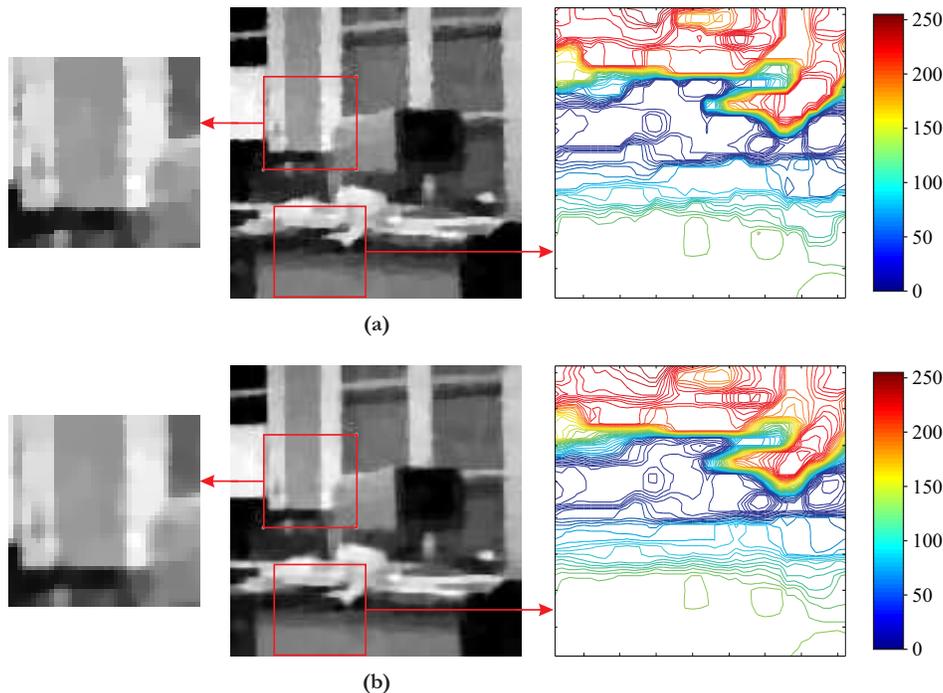


FIGURE 5. Convex reconstruction of the “office” image: (a) without wavelet-domain smoothing (MSE = 208.0, ISNR = 5.70 dB); (b) with wavelet-domain smoothing (MSE = 180.4, ISNR = 6.32 dB).

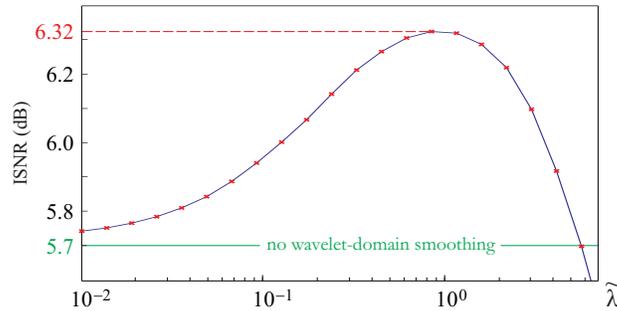


FIGURE 6. Convex reconstruction of the “office” image: ISNR as a function of the wavelet-domain smoothing parameter $\tilde{\lambda}$ (λ is set to its optimal value in terms of ISNR without wavelet-domain smoothing).

The above results clearly show that first-order spatial regularization benefits from wavelet-domain smoothing; but is the converse also true? To answer this question, Fig. 7 displays the best reconstruction obtained using wavelet-domain smoothing without spatial regularization (that is, $\lambda = 0$). This estimate is worse than that obtained using only spatial regularization (the “trellis-like” texture is due to the fact that wavelet-domain smoothing favors solutions with vertically-oriented horizontal subbands and horizontally-oriented vertical subbands). We can therefore conclude that spatial- and wavelet-domain smoothing work in synergy.

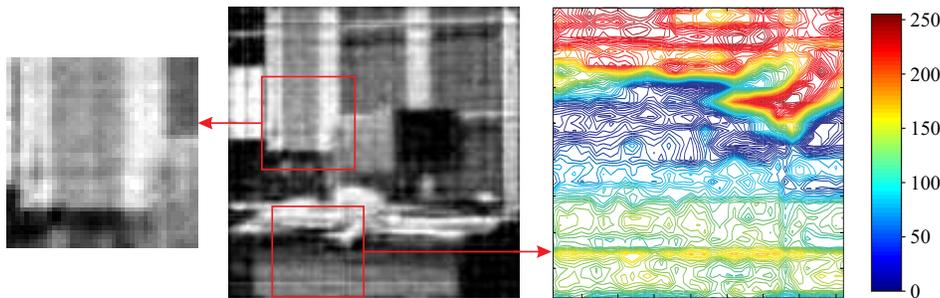


FIGURE 7. Reconstruction of the “office” image without spatial-domain smoothing (MSE = 375.1, ISNR = 3.14 dB).

5.1.2. *Non-convex case.* Fig. 8 shows the estimates obtained with the PF $\psi = \psi_{\text{HL}}^{10}$ for $(\lambda, \tilde{\lambda}) = (10.5, 0)$ and $(\lambda, \tilde{\lambda}) = (10.5, 2.3)$. The particular value $\lambda = 10.5$ was purposely selected to obtain a slightly under-regularized reconstruction in the absence of wavelet-domain smoothing (the optimal value is 21.0). The value $\tilde{\lambda} = 2.3$ is the optimal setting associated with $\lambda = 10.5$. The MSE and ISNR are 501.9 and 1.88 dB for the standard regularization approach, and 191.0 and 6.08 dB using wavelet-domain smoothing. The improvement is even more striking than in the convex case: the edge artifacts are almost completely removed, and the contour plots show substantial reduction of the staircase effect. As seen in Fig. 9, the evolution of the ISNR as a function of $\tilde{\lambda}$ is reasonably smooth, which indicates that

our optimization algorithm behaves well in the non-convex case. Although λ is not optimal (the best reconstruction achieved without wavelet-domain smoothing has an MSE of 354.1 and an ISNR of 3.40 dB), our model outperforms optimal standard regularization over a significant range of values of $\tilde{\lambda}$.

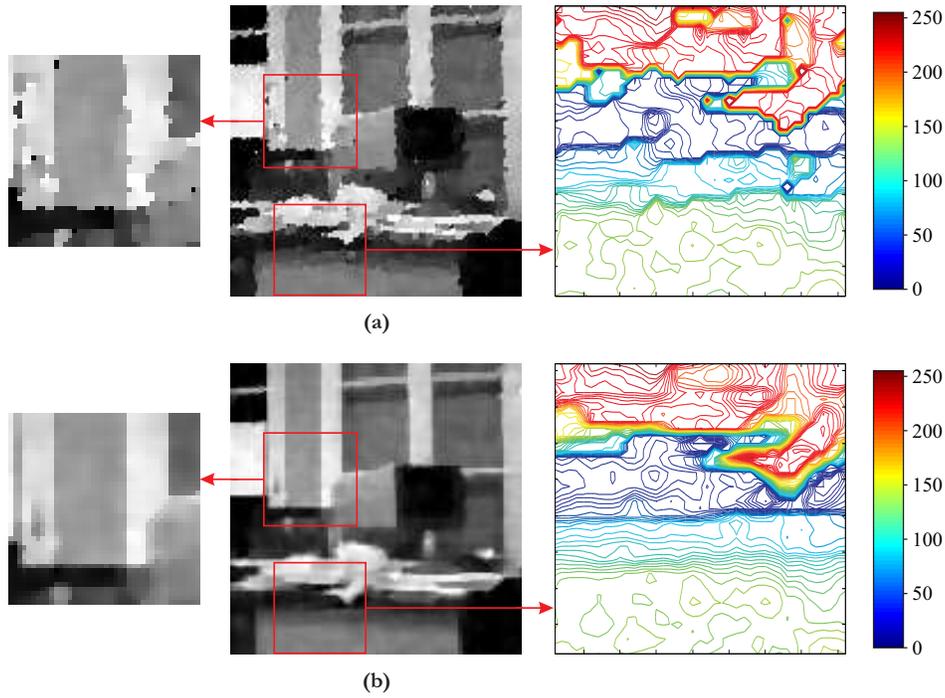


FIGURE 8. Non-convex reconstruction of the “office” image: (a) without wavelet-domain smoothing (MSE = 501.9, ISNR = 1.88 dB); (b) with wavelet-domain smoothing (MSE = 191.0, ISNR = 6.08 dB).

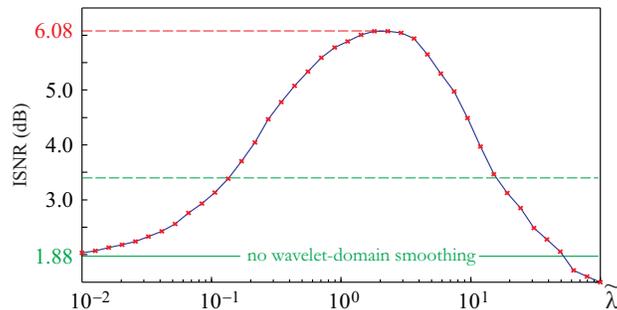


FIGURE 9. Non-convex reconstruction of the “office” image: ISNR as a function of the wavelet-domain smoothing parameter $\tilde{\lambda}$ (the dotted green line represents the highest ISNR achievable without wavelet-domain smoothing).

5.1.3. *Comparison with state-of-the-art methods.* Fig. 10 shows the estimates achieved by IFAS reconstruction, HFN regularization, and ITV2 regularization. (The HFN and ITV2 regularization estimates are the best obtained by adjusting the regularization strength.) ITV2 regularization produces the best solution among these three methods, but it fails to outperform first-order TV regularization: the ITV2 regularization estimate shown in Fig. 10(c) has an MSE of 211.2 and an ISNR of 5.64dB, whereas the first-order TV regularization estimate shown in Fig. 5(a) has an MSE of 191.0 and an ISNR of 5.70dB. This observation is consistent with the fact that the performances of second-order regularization techniques decline when the original image is too degraded. We also observe that HFN and ITV2 regularization produce estimates with a slightly granular appearance and tend to smooth edges (this unwanted smoothing effect is clearly visible from the enlargements of the cropped areas in Figs. 10(b) and 10(c)). As the comparison with Figs. 5(b) and 8(b) shows, the proposed combination of spatial- and wavelet-domain smoothing does not have these drawbacks—our approach performs significantly better in terms of MSE, ISNR, and visual quality.

5.1.4. *Comparison in terms of deviation from the original range.* Since the considered cost function does not constrain the pixel value range, the range of the computed solution \hat{x} may not be included in that of the original image x^* . In other words, there may exist some $k \in \llbracket 1, K \rrbracket$ such that $\hat{x}_k \notin [x_{\min}^*, x_{\max}^*]$, where x_{\min}^* and x_{\max}^* respectively denote the minimum and maximum pixel values in x^* . For proper comparison, the images shown in Figs. 5, 7, 8, and 10 are clipped versions of the actual solutions produced by the competing algorithms: the displayed image \check{x} associated with \hat{x} is defined by

$$(43) \quad \check{x}_k = \begin{cases} x_{\min}^* & \text{if } \hat{x}_k < x_{\min}^*, \\ \hat{x}_k & \text{if } \hat{x}_k \in [x_{\min}^*, x_{\max}^*], \\ x_{\max}^* & \text{if } \hat{x}_k > x_{\max}^*. \end{cases}$$

To measure the global deviation of the pixel values in \hat{x} from the range of x^* , we use a metric we call the *deviation from the original range* (DOR), which is defined by

$$(44) \quad \text{DOR} = \frac{\|\hat{x} - \check{x}\|_1}{K(x_{\max}^* - x_{\min}^*)}.$$

We consistently observe that the introduction of the wavelet-domain smoothing term does not introduce DOR if there is none and that it reduces the DOR otherwise. For instance, for the reconstruction of the “office” image, wavelet-domain smoothing reduces the DOR from $3.38 \cdot 10^{-4}$ to $3.08 \cdot 10^{-4}$ in the convex case, and from $1.79 \cdot 10^{-3}$ to $4.82 \cdot 10^{-4}$ in the non-convex case. The DOR values of the estimates obtained using wavelet-domain smoothing are also smaller than those of the solutions achieved by IFAS reconstruction, HFN regularization, and ITV2 regularization ($1.16 \cdot 10^{-3}$, $1.10 \cdot 10^{-3}$, and $8.58 \cdot 10^{-4}$, respectively).

5.1.5. *Comparison of intensity profiles.* Fig. 11 displays line intensity profiles of the estimates shown in Fig. 5 (convex case), Fig. 8 (non-convex case), and Figs. 10(b) and 10(c) (HFN and ITV2 regularization). We can see from Figs. 11(b) and 11(c) that the wavelet-domain smoothing term does not smooth the edges that are preserved by first-order spatial regularization. Furthermore, the addition of wavelet-domain smoothing allows to recover more details (see arrow 1) and eliminates offsets

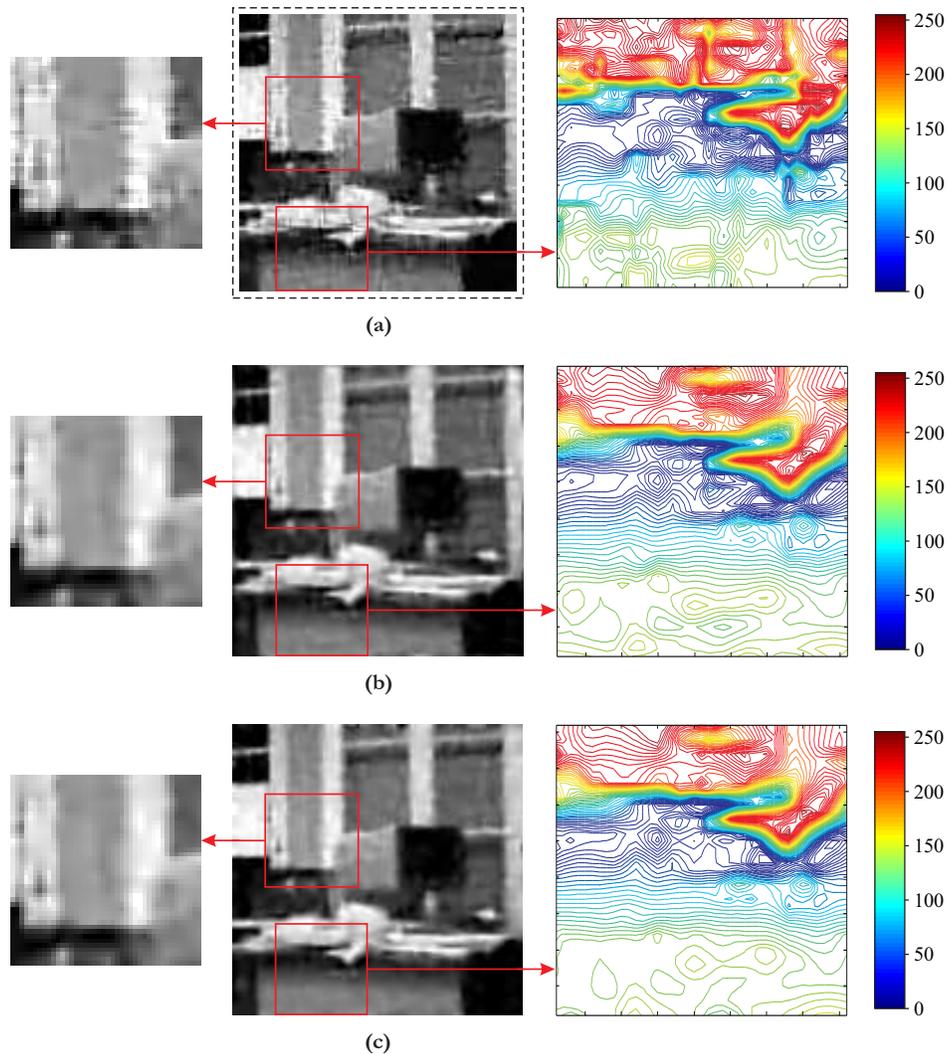


FIGURE 10. Reconstruction of the “office” image using state-of-the-art regularization approaches: (a) IFAS reconstruction (MSE = 282.7, ISNR = 4.37 dB); (b) HFN regularization (MSE = 222.0, ISNR = 5.42 dB); (c) ITV2 regularization (MSE = 211.2, ISNR = 5.64 dB).

in edge locations (see arrows 2 and 3). The edge-smoothing effect caused by HFN and ITV2 regularization is particularly visible in the regions pointed out by arrows 4 and 5.

5.2. “Peppers” image. The “peppers” image test-problem has the following objectives: (i) to examine the effect of the wavelet-domain smoothing term when the original image has reduced high-frequency content in the horizontal and vertical directions, (ii) to show that our algorithm behaves well when the data-fidelity term is not quadratic, and (iii) to capture the improvement brought by non-quadratic

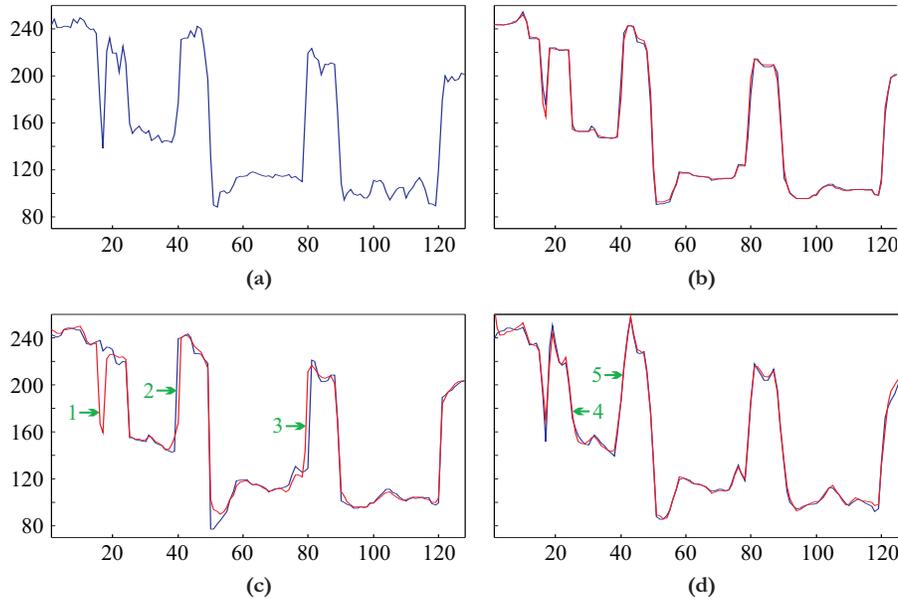


FIGURE 11. Intensity profiles along line 40 (displayed in cyan in Fig. 3(a)): (a) original; (b) convex reconstruction with (red) and without (blue) wavelet-domain smoothing; (c) non-convex reconstruction with (red) and without (blue) wavelet-domain smoothing; (d) HFN (blue) and ITV2 (red) regularization.

data-fidelity in the presence of impulse noise. Note that, contrary to the approaches proposed in [7, 8, 34], we do not remove the outlier pixels from the data before reconstruction.

5.2.1. *Quadratic versus non-quadratic data-fidelity.* Figs. 12(a) and 13(a) show the best reconstructions obtained using the square data-penalty function $\phi(t) = t^2$ and without wavelet-domain smoothing. They are qualitatively similar: the MSE and the ISNR are 559.3 and 6.16 dB in the convex case, and 568.4 and 6.09 dB in the non-convex case. Both estimates are over-smoothed, but decreasing the value of λ leads to worse solutions that are substantially biased by the impulse noise. By contrast, Figs. 12(b) and 13(b) show the best reconstructions obtained using the non-quadratic data-penalty function $\phi = \psi_{\text{id}}^{0.1}$. These solutions have much better contrast and sharpness than those achieved with quadratic data-fidelity. Quantitatively, changing quadratic data-fidelity for non-quadratic data-fidelity divides the MSE by about 4.5 in the convex-case and about 2.8 in the non-convex-case; the corresponding increases of the ISNR are 6.58 dB and 4.51 dB, respectively.

5.2.2. *Convex versus non-convex regularization.* Looking at the contour plots in Figs. 12(b) and 13(b), we observe that the edges obtained with the non-convex PF $\psi = \psi_{\text{HL}}^{10}$ are sharper than those obtained with the convex PF $\psi = \psi_{\text{id}}^{0.1}$. It should be stressed, however, that sharp-edge recovery makes complete sense when a significant proportion of the discontinuities in the original image are only one pixel wide—it can be a disadvantage otherwise. Indeed, non-convex regularization tends to create sharp discontinuities wherever the gradient exceeds a certain threshold, thus turning

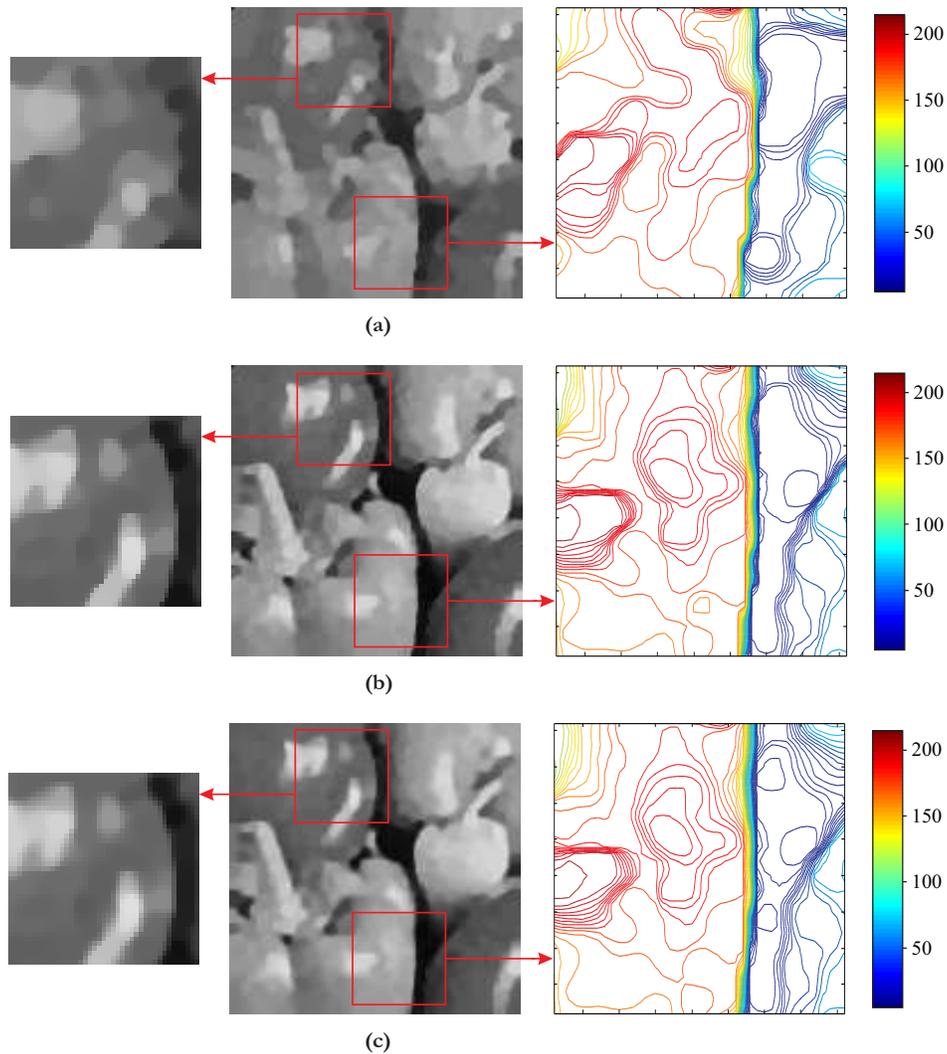


FIGURE 12. Convex reconstruction of the “peppers” image: (a) quadratic data-fidelity, no wavelet-domain smoothing (MSE = 559.3, ISNR = 6.16 dB); (b) non-quadratic data-fidelity, no wavelet-domain smoothing (MSE = 122.9, ISNR = 12.74 dB); (c) non-quadratic data-fidelity + wavelet-domain smoothing (MSE = 116.0, ISNR = 13.0 dB).

edges that are more than one pixel wide into thin edges that do not conform to reality. This unwanted effect is clearly observed by comparing the original contours in Fig. 4(a) and the contours in Fig. 13(b).

5.2.3. *Reconstruction with wavelet-domain smoothing.* Figs. 12(c) and 13(c) show the best solutions achieved by adding the wavelet-domain smoothing term (the values of λ are kept unchanged). In the convex case, the MSE decreases from 122.9 to 116.0, and the ISNR increases from 12.74 dB to 13.0 dB. Except for a slight

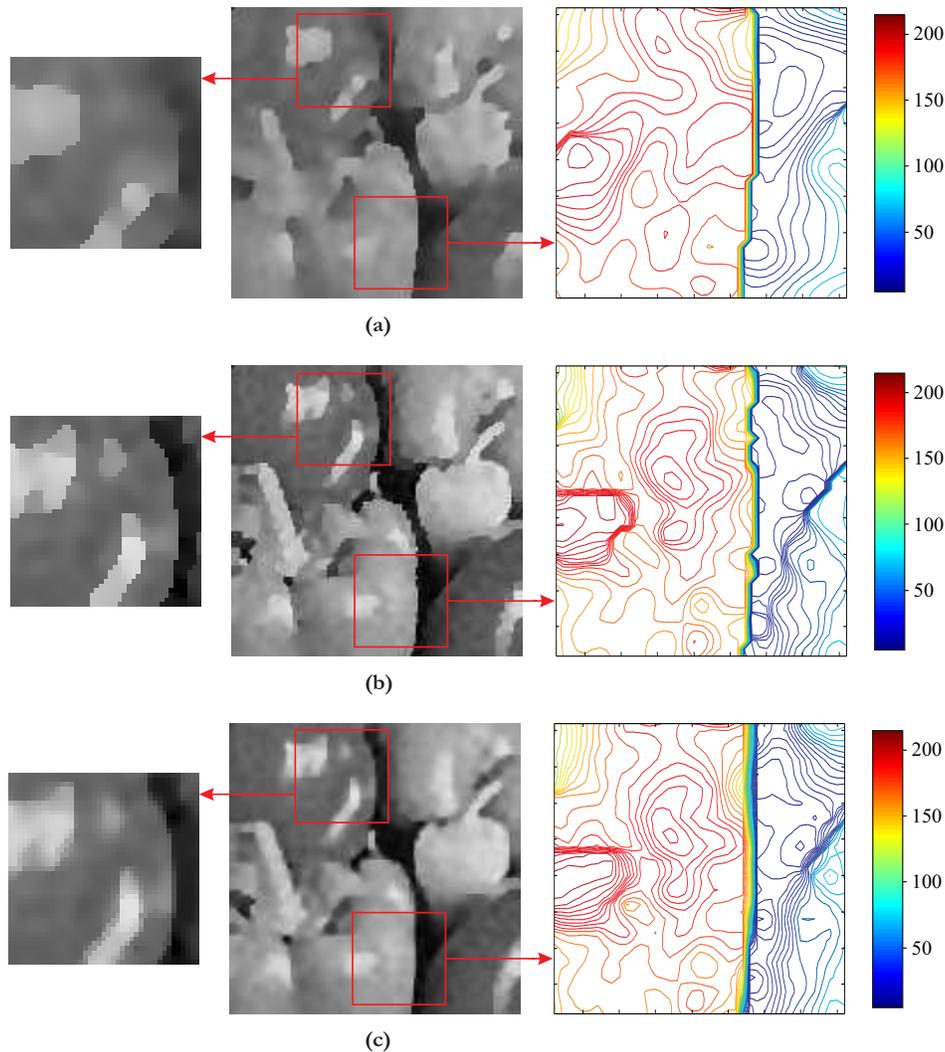


FIGURE 13. Non-convex reconstruction of the “peppers” image: (a) quadratic data-fidelity, no wavelet-domain smoothing (MSE = 568.4, ISNR = 6.09 dB); (b) non-quadratic data-fidelity, no wavelet-domain smoothing (MSE = 201.2, ISNR = 10.60 dB); (c) non-quadratic data-fidelity + wavelet-domain smoothing (MSE = 129.3, ISNR = 12.52 dB).

reduction of the staircase effect, there is not much difference with the solution obtained without wavelet-domain smoothing. This result is actually satisfactory: the wavelet-domain smoothing term has negligible effect if the reconstruction obtained without using it does not have noisy contour lines and does not present a staircase effect. In the non-convex case, the improvement accompanying the introduction of the wavelet-domain smoothing term is significant: the MSE decreases from 201.2 to 129.3, and the ISNR increases from 10.60 dB to 12.52 dB. As observed for the “office” image, the staircase effect is largely reduced and the noise on the contour

lines is eliminated. In addition, we notice that the wavelet-domain smoothing term allows to better recover the true thickness of the edges. This can be seen from the contour plots in Figs. 4(a), 13(b), and 13(c): the vertical edge obtained without wavelet-domain smoothing is too sharp compared to the original one, whereas that obtained using wavelet-domain smoothing is more conform to reality.

5.2.4. *Comparison with state-of-the-art methods.* Contrary to IFAS reconstruction, the algorithms proposed in [32, 28] for HFN and ITV2 regularization are limited to quadratic data-fidelity. Nevertheless, as mentioned in Section 2.1, the HFN and ITV2 regularizers are special cases of (4), and thus we can use our algorithm \mathbb{A}_V to assess their behavior when added to a non-quadratic data-fidelity term. Fig. 14 displays the estimates achieved by IFAS reconstruction and by the HFN and ITV2 regularizers combined with non-quadratic data-fidelity. Similarly to what observed for the “office” image, IFAS reconstruction gives the worse estimate, and HFN and ITV2 regularization are outperformed by first-order TV regularization. Again, our approach outperforms all three alternatives in terms of MSE and ISNR. We also clearly see from the enlargements and the contour plots that HFN and ITV2 regularization produce over-smoothed solutions compared to reconstruction with wavelet-domain smoothing.

6. Conclusion. We introduced the idea of *implicitly interacting discontinuities* by means of an additional penalty term operating on the detail images in the wavelet domain. When combined with first-order edge-preserving regularization, the proposed wavelet-domain regularizer preserves boundary sharpness while smoothing contour lines. Aside from producing visually more pleasing reconstructions, this behavior is desirable for subsequent feature-extraction and segmentation tasks. We also provided an efficient half-quadratic optimization algorithm with strong convergence guarantees even when the faithfulness to the data is non-quadratic and the cost function is not convex. Our experiments show that our approach outperforms first-order TV regularization as well as state-of-the-art second-order regularization techniques. The only drawback is the introduction of the additional hyper-parameter $\tilde{\lambda}$ which controls the degree of smoothing in the wavelet domain. However, the range of values of $\tilde{\lambda}$ that lead to an improvement in reconstruction quality is fairly large, and preliminary results suggest that $\tilde{\lambda}$ can be selected by tracking the value of the spatial-domain regularizer Ψ when this parameter varies.

Appendix A. Proof of Theorem 4.1.

A.1. **General framework.** We consider a generalization of the optimization problems studied in Section 4, namely, the minimization of a cost function $Z : \mathbb{R}^K \rightarrow \mathbb{R}_+$ of the form

$$(45) \quad Z(x) = F_0(x) + \sum_{i \in \llbracket 1, r \rrbracket} \lambda_i (f_i \circ F_i)(x),$$

where $\lambda_i > 0$ for all i , and where the functions $F_i : \mathbb{R}^K \rightarrow \mathbb{R}_+$ and $f_i : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfy the following assumptions:

A1. For any $i \in \llbracket 0, r \rrbracket$, F_i is differentiable.

A2. For any $i \in \llbracket 1, r \rrbracket$, f_i is twice differentiable, strictly concave, and such that

$$f_i(0) = 0, \quad f_i'(0) = 1, \quad \text{and} \quad \lim_{t \rightarrow +\infty} f_i'(t) = 0.$$

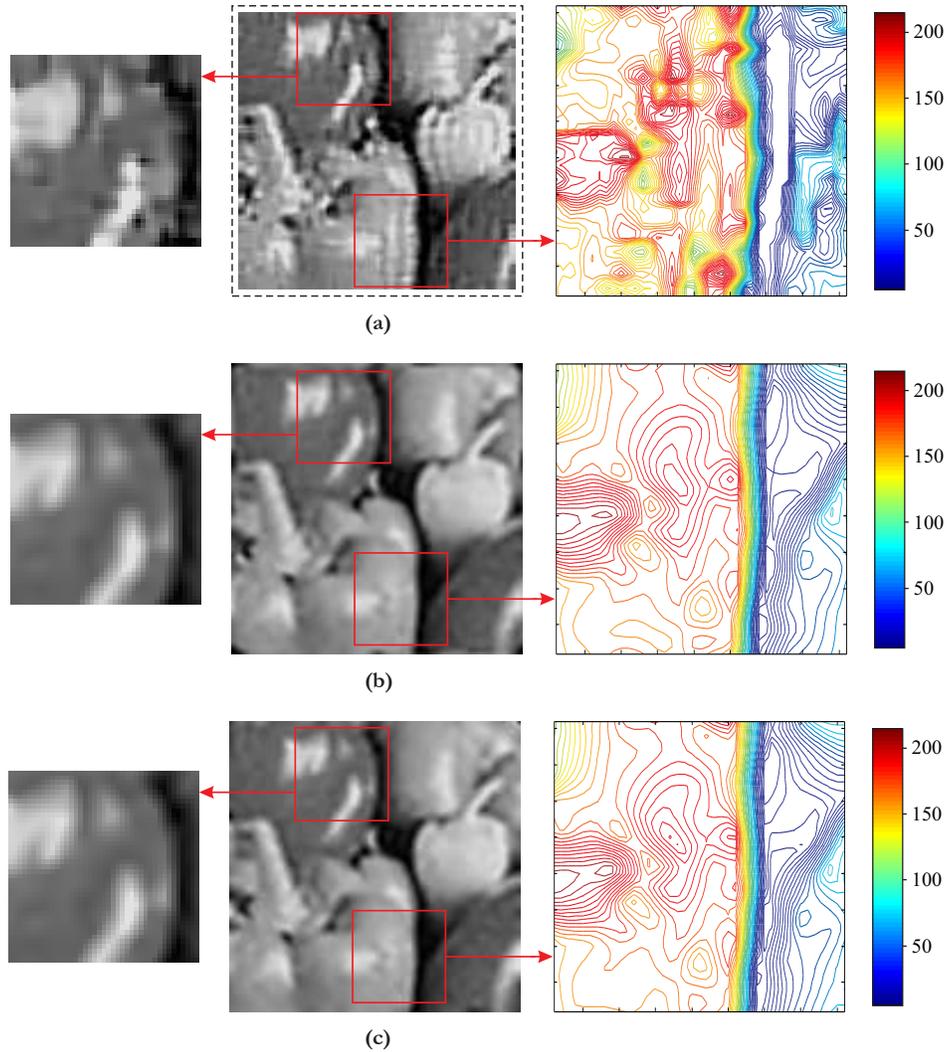


FIGURE 14. Reconstruction of the “peppers” image using state-of-the-art regularization approaches: (a) IFAS reconstruction (MSE = 376.4, ISNR = 7.88 dB); (b) HFN regularization combined with non-quadratic data-fidelity (MSE = 135.5, ISNR = 12.32 dB); (c) ITV2 regularization combined with non-quadratic data-fidelity (MSE = 135.3, ISNR = 12.32 dB).

A3. Z is coercive, that is, $\lim_{\|x\| \rightarrow \infty} Z(x) = +\infty$.

We also assume that the function $Z_0 : \mathbb{R}^K \times (0, 1]^r \rightarrow \mathbb{R}_+$ defined by

$$(46) \quad Z_0(x, \gamma) = F_0(x) + \sum_{i \in [1, r]} \lambda_i \gamma_i F_i(x)$$

is such that

A4. for any $\gamma \in (0, 1]^r$, $Z_0(\cdot, \gamma) : x \in \mathbb{R}^K \mapsto Z_0(x, \gamma)$ is strictly convex.

Let $g = (g_1, \dots, g_r) : \mathbb{R}^K \rightarrow \mathbb{R}^r$ be the vector-valued function with components

$$(47) \quad g_i = f'_i \circ F_i,$$

and let $\mathcal{J} : \mathbb{R}^K \rightarrow \mathbb{R}^r$ be given by

$$(48) \quad \mathcal{J}(y) = \arg \inf_{x \in \mathbb{R}^K} Z_0(x, g(y)).$$

(The function \mathcal{J} is well defined. Indeed, for any $\gamma \in (0, 1]^r$, we have

$$(49) \quad \forall x \in \mathbb{R}^K, \quad \frac{Z_0(x, \gamma)}{\inf_{i \in \llbracket 1, r \rrbracket} \gamma_i} \geq Z_0(x, \mathbf{1}_r) \geq Z(x),$$

where $\mathbf{1}_r = (1, \dots, 1) \in \mathbb{R}^r$, and where the right inequality follows from A2. Therefore, $Z_0(\cdot, \gamma)$ is coercive by A3, and since it is strictly convex, it has a unique global minimum.) We will show in Appendix A.3 that the cost functions of type (45) that satisfy assumptions A1–A4 are minimized by the algorithm

$$(50) \quad \mathbb{A}_Z \equiv \begin{cases} x^{(0)} \in \mathbb{R}^K, \\ \forall n \in \mathbb{N}, \quad x^{(n+1)} = \mathcal{J}(x^{(n)}). \end{cases}$$

But first, we show in Appendix A.2 that the cost functions and the optimization algorithms studied in Section 4 are special cases of this general framework — Theorem 4.1 then readily follows from the convergence results stated in Theorem A.9 in Appendix A.3.

A.2. Special cases of interest.

Theorem A.1. *If conditions C1–C3 are satisfied, then the cost function W (35) fits the general form (45) along with assumptions A1–A4, and the algorithm \mathbb{A}_W (37) is a special case of \mathbb{A}_Z . (The same conclusion holds for the cost function V (22) and the algorithm \mathbb{A}_V (30), as these are special cases of W and \mathbb{A}_W .)*

Proof. Let

$$I_\theta = \llbracket 1, L' \rrbracket, \quad I_\phi = \begin{cases} \llbracket L'+1, L'+K' \rrbracket & \text{if C1.1,} \\ \emptyset & \text{if C1.2,} \end{cases}$$

$$\text{and } r = |I_\theta \cup I_\phi| = \begin{cases} L' + K' & \text{if C1.1,} \\ L' & \text{if C1.2.} \end{cases}$$

We consider the function Z of the form (45) with r given above and with

$$(51) \quad F_0(x) = \begin{cases} \mu \|(b - \mathcal{A}x)^+\|_2^2 & \text{if C1.1,} \\ \|\mathcal{H}x - d\|_2^2 + \mu \|(b - \mathcal{A}x)^+\|_2^2 & \text{if C1.2,} \end{cases}$$

$$(52) \quad \forall i \in I_\theta, \quad \begin{cases} \lambda_i = \theta''_i(0)/2, \\ f_i(t) = \theta_i(\sqrt{t})/\lambda_i, \\ F_i(x) = \|\mathcal{Q}_i x\|_2^2, \end{cases}$$

$$(53) \quad \text{and } \forall i \in I_\phi, \quad \begin{cases} \lambda_i = \phi''_{i-L'}(0)/2, \\ f_i(t) = \phi_{i-L'}(\sqrt{t})/\lambda_i, \\ F_i(x) = (\mathcal{H}x - d)_{i-L'}^2. \end{cases}$$

Clearly, $Z = W$, and assumption A1 is satisfied. It remains to check the validity of A2–A4 and the equivalence of \mathbb{A}_Z and \mathbb{A}_W .

Assumption A2. Each f_i is of the form $f_i(t) = 2\varphi(\sqrt{t})/\varphi''(0)$ with $\varphi \in \mathfrak{S}$. We immediately have that $f_i(0) = 0$, f'_i is strictly decreasing on $(0, +\infty)$, $\lim_{t \rightarrow 0^+} f'_i(t) = 1$, $\lim_{t \rightarrow +\infty} f'_i(t) = 0$, and f_i is twice differentiable on $(0, +\infty)$. Without going into details, the second-order Taylor expansion of φ at zero gives that $f'_i(0) = 1$ (hence f'_i is strictly decreasing on \mathbb{R}_+ , and thus f_i is strictly concave), and the third-order Taylor expansion of φ' at zero gives that f_i is twice differentiable at zero.

Assumption A3. Suppose that V is not coercive, that is, there exists a sequence $(x^{(n)})_n$ in \mathbb{R}^K such that $\lim_{n \rightarrow \infty} \|x^{(n)}\| = \infty$ and $(V(x^{(n)}))_n$ is bounded. Since the ϕ_k 's and the θ_l 's are non-negative, and since $\lim_{t \rightarrow +\infty} \phi_k(t) = \lim_{t \rightarrow +\infty} \theta_l(t) = +\infty$ for all k and for all l , $(V(x^{(n)}))_n$ is bounded if and only if the sequences $(\sup_k |(\mathcal{H}x^{(n)} - d)_k|)_n$ and $(\sup_l \|\mathcal{Q}_l x^{(n)}\|_2)_n$ are bounded, which is equivalent to say that $(\|\mathcal{H}x^{(n)}\|)_n$ and $(\|\mathcal{Q}x^{(n)}\|)_n$ are bounded. Let \mathfrak{q} be the quadratic form on \mathbb{R}^K defined by $\mathfrak{q}(x) = \|\mathcal{H}x\|^2 + \|\mathcal{Q}x\|^2$. By C3, \mathfrak{q} is positive-definite (or, equivalently, coercive), and thus $\lim_{n \rightarrow \infty} \mathfrak{q}(x^{(n)}) = +\infty$. This contradicts the fact that $(\|\mathcal{H}x^{(n)}\|)_n$ and $(\|\mathcal{Q}x^{(n)}\|)_n$ are bounded. Therefore, V is coercive, and thus so is W .

Assumption A4. For any $\gamma \in (0, 1]^r$, we let

$$\begin{aligned} \gamma^\theta &= (\gamma_1 \theta''_1(0), \dots, \gamma_{L'} \theta''_{L'}(0)), \\ \text{and } \gamma^\phi &= \begin{cases} (\gamma_{L'+1} \phi''_1(0), \dots, \gamma_{L'+K'} \phi''_{K'}(0)) & \text{if C1.1,} \\ 2 \cdot \mathbf{1}_{K'} & \text{if C1.2.} \end{cases} \end{aligned}$$

Substituting (51)–(53) into (46), we obtain

$$Z_0(x, \gamma) = \frac{1}{2} W_0(x, \gamma^\phi, \gamma^\theta).$$

Since $W_0(\cdot, \delta, \varepsilon)$ is strictly convex for any positive δ and ε , it follows that $Z_0(\cdot, \gamma)$ is strictly convex for any $\gamma \in (0, 1]^r$.

Equivalence of \mathbb{A}_Z and \mathbb{A}_W . Substituting (52) and (53) into (47), we get

$$g_i(y) = \begin{cases} \frac{\theta_i^\dagger(\|\mathcal{Q}_i y\|_2)}{\theta_i''(0)} & \text{if } i \in I_\theta, \\ \frac{\phi_{i-L'}^\dagger(|(\mathcal{H}y - d)_{i-L'}|)}{\phi_{i-L'}''(0)} & \text{if } i \in I_\phi. \end{cases}$$

Therefore,

$$\gamma = g(y) \iff (\gamma^\phi, \gamma^\theta) = (e_\phi(y), e_\theta(y)),$$

and thus

$$Z_0(x, g(y)) = \frac{1}{2} W_0(x, e_\phi(y), e_\theta(y)).$$

This shows that \mathbb{A}_Z and \mathbb{A}_W are equivalent algorithms. □

A.3. Convergence analysis of \mathbb{A}_Z . In this section, we consider the general framework described in Appendix A.1, and we assume that assumptions A1–A4 hold. We let $\mathfrak{C}(Z)$ be the set of critical points of Z and $\mathfrak{F}(\mathcal{J})$ be the set of fixed points of \mathcal{J} , that is,

$$(54) \quad \mathfrak{C}(Z) = \{x \in \mathbb{R}^K \mid \nabla Z(x) = 0\},$$

$$(55) \quad \text{and } \mathfrak{F}(\mathcal{J}) = \{x \in \mathbb{R}^K \mid \mathcal{J}(x) = x\}.$$

We denote by $\mathfrak{A}(x^{(n)})$ the set of accumulation points of a sequence $(x^{(n)})_n$.

We first show in Lemma A.2 that finding a critical point of Z is equivalent to finding a fixed point of \mathcal{J} . We will then use specific instances of Theorems 3.1 and 3.5 in [39] to establish convergence to a fixed point of \mathcal{J} —the results that we need are put together into Theorem A.4 after some definitions. The purpose of the subsequent lemmas (A.5 to A.8) is to show that Theorem A.4 applies when its variables ζ and \mathcal{K} are restrictions of Z and \mathcal{J} to compact sets of the form $\{x \in \mathbb{R}^K \mid Z(x) \leq a\}$, $a \in (\inf_{x \in \mathbb{R}^K} Z(x), +\infty)$. Our main results about the convergence of \mathbb{A}_Z are stated in Theorem A.9.

Lemma A.2. $\mathfrak{F}(\mathcal{J}) = \mathfrak{C}(Z)$.

Proof. Let $x \in \mathbb{R}^K$. We have

$$\nabla Z(x) = \nabla F_0(x) + \sum_{i \in [1, r]} \lambda_i g_i(x) \nabla F_i(x) = \nabla Z_0(\cdot, g(x))(x),$$

where $\nabla Z_0(\cdot, \gamma)(x)$ denotes the gradient of $Z_0(\cdot, \gamma) : y \in \mathbb{R}^K \mapsto Z_0(y, \gamma)$ evaluated at x . Therefore,

$$x \in \mathfrak{C}(Z) \iff \nabla Z_0(\cdot, g(x))(x) = 0 \iff x \in \mathfrak{F}(\mathcal{J}).$$

(The second equivalence follows from the fact that $Z_0(\cdot, \gamma)$ is strictly convex and coercive for any $\gamma \in (0, 1]^r$.) \square

Definitions A.3. Let Ω be a closed subset of \mathbb{R}^K , let \mathcal{K} be a function from Ω to Ω , and let $\mathfrak{F}(\mathcal{K})$ be the set of fixed points of \mathcal{K} .

(i) A sequence $(x^{(n)})_n$ in Ω is said to be generated by \mathcal{K} if

$$(56) \quad \begin{cases} x^{(0)} \in \Omega, \\ \forall n \in \mathbb{N}, x^{(n+1)} = \mathcal{K}(x^{(n)}). \end{cases}$$

(ii) \mathcal{K} is said to be uniformly compact if there exists a compact subset C of \mathbb{R}^K such that $\mathcal{K}(\Omega) \subseteq C$.

(iii) \mathcal{K} is said to be strictly monotonic with respect to a function $\zeta : \Omega \rightarrow \mathbb{R}$ if

$$(57) \quad \forall x \in \Omega \setminus \mathfrak{F}(\mathcal{K}), (\zeta \circ \mathcal{K})(x) < \zeta(x).$$

(iv) \mathcal{K} is said to be upper semi-continuous if, for any convergent sequences $(x^{(n)})_n$ and $(y^{(n)})_n$ in Ω ,

$$(58) \quad (\forall n \in \mathbb{N}, y^{(n)} = \mathcal{K}(x^{(n)})) \implies \lim_{n \rightarrow \infty} y^{(n)} = \mathcal{K}\left(\lim_{n \rightarrow \infty} x^{(n)}\right).$$

Theorem A.4. Let Ω be a closed subset of \mathbb{R}^K , let $\zeta : \Omega \rightarrow \mathbb{R}$ be a continuous function, and let $\mathcal{K} : \Omega \rightarrow \Omega$ be a function that is uniformly compact, strictly monotonic with respect to ζ , and upper semi-continuous. Let $(x^{(n)})_n$ be any sequence generated by \mathcal{K} .

(i) $\mathfrak{A}(x^{(n)}) \subseteq \mathfrak{F}(\mathcal{K})$.

(ii) There exists $x^* \in \mathfrak{F}(\mathcal{K})$ such that $\lim_{n \rightarrow \infty} \zeta(x^{(n)}) = \zeta(x^*)$.

(iii) Either $(x^{(n)})_n$ converges, or $\mathfrak{A}(x^{(n)})$ is a continuum.

(iv) Let x^* be an isolated fixed point of \mathcal{K} . If x^* is a strict local minimum of ζ , then there exists an open neighborhood \mathfrak{N}^* of x^* such that $\lim_{n \rightarrow \infty} x^{(n)} = x^*$ whenever $x^{(0)} \in \mathfrak{N}^*$.

Lemma A.5. For any $i \in \llbracket 1, r \rrbracket$, there exists a strictly increasing and strictly convex function $h_i : (0, 1] \rightarrow \mathbb{R}_+$ such that, for any $t \in \mathbb{R}_+$,

$$(59) \quad f_i(t) = \inf_{u \in (0, 1]} (tu + h_i(u)),$$

$$(60) \quad \text{and} \quad \arg \inf_{u \in (0, 1]} (tu + h_i(u)) = \{f'_i(t)\}.$$

Proof. Let $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a function satisfying the conditions of assumption A2 (we drop the subscript i to simplify notation). Then, f has the following properties:

$$(61) \quad \forall (s, t) \in (\mathbb{R}_+)^2, \quad s \neq t \implies f(t) < f(s) + (t - s)f'(s),$$

f'' is negative on $(0, +\infty)$, f' is a bijection from \mathbb{R}_+ to $(0, 1]$, and $(f')^{-1}$ is positive on $(0, 1)$.

Let $t \in \mathbb{R}_+$. We have from (61) that

$$f(t) = \inf_{s \in \mathbb{R}_+} (f(s) + (t - s)f'(s)).$$

For any $s \in \mathbb{R}_+$, there is a unique $u_s \in (0, 1]$ such that $u_s = f'(s)$ (or, equivalently, $s = (f')^{-1}(u_s)$), and thus we can write

$$f(s) + (t - s)f'(s) = tu_s + h(u_s),$$

where $h : (0, 1] \rightarrow \mathbb{R}$ is given by

$$h(u) = (f \circ (f')^{-1})(u) - (f')^{-1}(u)u.$$

Therefore,

$$(62) \quad f(t) = \inf_{s \in \mathbb{R}_+} (tu_s + h(u_s)) = \inf_{u \in (0, 1]} (tu + h(u)).$$

The function h is twice differentiable on $(0, 1)$. For any $u \in (0, 1)$, we have

$$h'(u) = -(f')^{-1}(u) < 0$$

$$\text{and} \quad h''(u) = \frac{-1}{(f'' \circ (f')^{-1})(u)} > 0.$$

Hence, h is strictly decreasing and strictly convex on $(0, 1)$. Since, in addition, h is continuous and $h(1) = 0$, it follows that h is non-negative, strictly decreasing, and strictly convex.

Let $t \in \mathbb{R}_+$. We have $(h \circ f')(t) = f(t) - tf'(t)$ (by the definition of h), and thus (62) gives

$$tf'(t) + (h \circ f')(t) = \inf_{u \in (0, 1]} (tu + h(u)).$$

The map $u \in (0, 1] \mapsto tu + h(u)$ is strictly convex (as h is strictly convex), and thus it has at most one minimum. Consequently, $\arg \inf_{u \in (0, 1]} (tu + h(u)) = \{f'(t)\}$. \square

Lemma A.6. Let $Z^* : \mathbb{R}^K \times (0, 1]^r \rightarrow \mathbb{R}_+$ be the dual cost function defined by

$$(63) \quad Z^*(x, \gamma) = F_0(x) + \sum_{i \in \llbracket 1, r \rrbracket} \lambda_i (\gamma_i F_i(x) + h_i(\gamma_i)),$$

where the h_i 's are the functions introduced in Lemma A.5. Then, for any $x \in \mathbb{R}^K$,

$$(64) \quad Z(x) = \inf_{\gamma \in (0, 1]^r} Z^*(x, \gamma) = Z^*(x, g(x)).$$

Proof. Let $x \in \mathbb{R}^K$. Using Lemma A.5, we have

$$\begin{aligned} Z(x) &= F_0(x) + \sum_{i \in \llbracket 1, r \rrbracket} \lambda_i \inf_{\gamma_i \in (0, 1]} (F_i(x)\gamma_i + h_i(\gamma_i)) \\ &= F_0(x) + \inf_{(\gamma_1, \dots, \gamma_r) \in (0, 1]^r} \sum_{i \in \llbracket 1, r \rrbracket} \lambda_i (F_i(x)\gamma_i + h_i(\gamma_i)) \\ &= \inf_{\gamma \in (0, 1]^r} Z^*(x, \gamma), \end{aligned}$$

and

$$Z(x) = F_0(x) + \sum_{i \in \llbracket 1, r \rrbracket} \lambda_i (F_i(x) \underbrace{(f'_i \circ F_i)}_{g_i}(x) + (h_i \circ \underbrace{f'_i \circ F_i}_{g_i})(x)) = Z^*(x, g(x)).$$

□

Lemma A.7. *The function \mathcal{J} is strictly monotonic with respect to Z , that is,*

$$(65) \quad \forall x \in \mathbb{R}^K \setminus \mathfrak{F}(\mathcal{J}), \quad (Z \circ \mathcal{J})(x) < Z(x).$$

Proof. Let $x \in \mathbb{R}^K$ such that $x \neq \mathcal{J}(x)$, and let $y = \mathcal{J}(x)$. We have

$$Z_0(y, g(x)) = \inf_{z \in \mathbb{R}^K} Z_0(z, g(x)) < Z_0(x, g(x)),$$

and thus the dual cost function Z^* (63) satisfies

$$Z^*(y, g(x)) - Z^*(x, g(x)) = Z_0(y, g(x)) - Z_0(x, g(x)) < 0.$$

Besides, by Lemma A.6,

$$Z(y) - Z(x) = Z^*(y, g(y)) - Z^*(x, g(x)) \leq Z^*(y, g(x)) - Z^*(x, g(x)),$$

and consequently $Z(y) < Z(x)$. □

Lemma A.8. *Let A be a bounded subset of \mathbb{R}^K , and let*

$$(66) \quad \Omega(A) = \{x \in \mathbb{R}^K \mid Z(x) \leq \sup Z(A)\}.$$

- (i) $\Omega(A)$ is compact.
- (ii) $\mathcal{J}(\Omega(A)) \subseteq \Omega(A)$.
- (iii) $\mathcal{J}|_{\Omega(A)}$ is upper semi-continuous.
- (iv) Let $(x^{(n)})_n$ be a sequence generated by \mathbb{A}_Z , and let $\Omega_0 = \Omega(\{x^{(0)}\})$ and $\Omega_1 = \Omega(\{x^{(1)}\})$. If $x^{(0)} \notin \mathfrak{F}(\mathcal{J})$, then

$$(67) \quad \forall n \in \mathbb{N}^*, \quad x^{(n)} \in \Omega_1 \subset \Omega_0^\circ,$$

where Ω_0° denotes the interior of Ω_0 .

Proof. (i) The set $\Omega(A) = Z^{-1}([0, \sup Z(A)])$ is closed because Z is continuous. Besides, since Z is coercive,

$$\exists a > 0, \quad \forall x \in \mathbb{R}^K, \quad \|x\| > a \implies Z(x) > \sup Z(A),$$

that is, there exists $a > 0$ such that $\Omega(A) \subseteq \{x \in \mathbb{R}^K \mid \|x\| \leq a\}$. Therefore, $\Omega(A)$ is a closed bounded subset of \mathbb{R}^K .

(ii) Let $x \in \Omega(A)$. Using Lemma A.7, we have

$$(Z \circ \mathcal{J})(x) \leq Z(x) \leq \sup Z(A),$$

and thus $\mathcal{J}(x) \in \Omega(A)$.

(iii) Let $(x^{(n)})_n$ and $(y^{(n)})_n$ be two convergent sequences in $\Omega(A)$ such that $y^{(n)} = \mathcal{J}(x^{(n)})$ for all n , and let $x^* = \lim_{n \rightarrow \infty} x^{(n)}$ and $y^* = \lim_{n \rightarrow \infty} y^{(n)}$. Let $z \in \mathbb{R}^K$. By the definition of \mathcal{J} , we have

$$(68) \quad Z_0(y^{(n)}, g(x^{(n)})) \leq Z_0(z, g(x^{(n)}))$$

for all n . Since Z_0 and g are continuous, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} Z_0(y^{(n)}, g(x^{(n)})) &= Z_0(y^*, g(x^*)) \\ \text{and } \lim_{n \rightarrow \infty} Z_0(z, g(x^{(n)})) &= Z_0(z, g(x^*)). \end{aligned}$$

Therefore, passing to the limit in (68), we obtain

$$Z_0(y^*, g(x^*)) \leq Z_0(z, g(x^*)).$$

This shows that $y^* = \arg \inf_{z \in \mathbb{R}^K} Z_0(z, g(x^*))$ (as $Z_0(\cdot, g(x^*))$ is strictly convex), that is, $y^* = \mathcal{J}(x^*)$.

(iv) Let $(x^{(n)})_n$ be a sequence generated by \mathbb{A}_Z such that $x^{(0)} \notin \mathfrak{F}(\mathcal{J})$. By Lemma A.7, we have $Z(x^{(1)}) < Z(x^{(0)})$ and $Z(x^{(n+1)}) \leq Z(x^{(n)})$ for all $n \in \mathbb{N}^*$. Consequently, $Z(x^{(n)}) \leq Z(x^{(1)})$ (or, equivalently, $x^{(n)} \in \Omega_1$) for all $n \in \mathbb{N}^*$, and

$$\Omega_1 \subset \{x \in \mathbb{R}^K \mid Z(x) < Z(x^{(0)})\} \subseteq \Omega_0^\circ.$$

□

Theorem A.9. *Let $(x^{(n)})_n$ be any sequence generated by \mathbb{A}_Z under assumptions A1–A4.*

- (I) $\emptyset \neq \mathfrak{A}(x^{(n)}) \subseteq \mathfrak{C}(Z)$.
- (II) *There exists $x^* \in \mathfrak{C}(Z)$ such that $\lim_{n \rightarrow \infty} Z(x^{(n)}) = Z(x^*)$, and, for any $n \in \mathbb{N}$, $Z(x^{(n+1)}) \leq Z(x^{(n)})$ with equality if and only if $x^{(n)} \in \mathfrak{C}(Z)$.*
- (III) *Either $(x^{(n)})_n$ converges, or $\mathfrak{A}(x^{(n)})$ is a continuum.*
- (IV) *Let x^* be an isolated critical point of Z . If x^* is a local minimum of Z , then there exists an open neighborhood \mathfrak{N} of x^* such that $\lim_{n \rightarrow \infty} x^{(n)} = x^*$ whenever $x^{(0)} \in \mathfrak{N}$.*
- (V) $\lim_{n \rightarrow \infty} \inf_{y \in \mathfrak{C}(Z)} \|x^{(n)} - y\| = 0$.
- (VI) *If $\mathfrak{C}(Z)$ is discrete, then there exists $x^* \in \mathfrak{C}(Z)$ such that $\lim_{n \rightarrow \infty} x^{(n)} = x^*$.*
- (VII) *If Z is strictly convex, then $(x^{(n)})_n$ converges to the global minimum of Z .*

Proof. (I)–(III) By Lemmas A.7 and A.8, $(x^{(n)})_n$ is generated by $\mathcal{J}|_{\Omega_0}$ (in the sense of Definition A.3(i)), and the assumptions of Theorem A.4 hold when $\Omega = \Omega_0$, $\mathcal{K} = \mathcal{J}|_{\Omega_0}$, and $\zeta = Z|_{\Omega_0}$. Therefore, since $\mathfrak{F}(\mathcal{J}|_{\Omega_0}) \subseteq \mathfrak{C}(Z)$ by Lemma A.2, we have the following results:

- $\mathfrak{A}(x^{(n)}) \subseteq \mathfrak{C}(Z)$.
- There exists $x^* \in \mathfrak{C}(Z)$ such that $\lim_{n \rightarrow \infty} Z(x^{(n)}) = Z(x^*)$.
- Either $(x^{(n)})_n$ converges, or $\mathfrak{A}(x^{(n)})$ is a continuum.

Since $(x^{(n)})_{n \geq 1}$ is in the compact set Ω_1 , it has a convergent subsequence, and thus $\mathfrak{A}(x^{(n)})$ is not empty. The strict monotonicity of \mathcal{J} gives that, for any $n \in \mathbb{N}$, $Z(x^{(n+1)}) \leq Z(x^{(n)})$ with equality if and only if $x \in \mathfrak{C}(Z)$.

(IV) Let x^* be an isolated critical point and a local minimum of Z . Then, x^* is a strict local minimum, because otherwise it is not an isolated critical point. Let $a > 0$, let $B(x^*, a)$ be an open ball around x^* with radius a , and let $\Omega_a = \Omega(B(x^*, a))$ be defined as in (66). By lemmas A.7 and A.8, the sequences generated by $\mathcal{J}|_{\Omega_a}$ are the sequences $(x^{(n)})_n$ generated by \mathbb{A}_Z such that $x^{(0)} \in \Omega_a$, and the assumptions of

Theorem A.4 hold when $\Omega = \Omega_a$, $\mathcal{K} = \mathcal{J}|_{\Omega_a}$, and $\zeta = Z|_{\Omega_a}$. Therefore, there exists an open neighborhood \mathfrak{N}^* of x^* such that, for any sequence $(x^{(n)})_n$ generated by \mathbb{A}_Z ,

$$x^{(0)} \in \mathfrak{N}^* \cap \Omega_a \implies \lim_{n \rightarrow \infty} x^{(n)} = x^*.$$

Hence, it suffices to take $\mathfrak{N} = \mathfrak{N}^* \cap \Omega_a^\circ$, where Ω_a° denotes the interior of Ω_a .

(v) Let $d_{\mathfrak{C}(Z)} : \mathbb{R}^K \rightarrow \mathbb{R}_+$ be defined by $d_{\mathfrak{C}(Z)}(x) = \inf_{y \in \mathfrak{C}(Z)} \|x - y\|$, and suppose that $d_{\mathfrak{C}(Z)}(x^{(n)})$ does not go to zero as $n \rightarrow \infty$; that is, there exists $a > 0$ such that

$$\forall m \in \mathbb{N}, \exists n \geq m, d_{\mathfrak{C}(Z)}(x^{(n)}) > a.$$

Then, there exists a subsequence $(y^{(l)})_l$ of $(x^{(n)})_{n \geq 1}$ such that $d_{\mathfrak{C}(Z)}(y^{(l)}) > a$ for all l , and since $(x^{(n)})_{n \geq 1}$ is in the compact set Ω_1 , $(y^{(l)})_l$ has a convergent subsequence $(z^{(k)})_k$. The limit point z^* of $(z^{(k)})_k$ is an accumulation point of $(x^{(n)})_n$, but $z^* \notin \mathfrak{C}(Z)$ (as $d_{\mathfrak{C}(Z)}(z^{(k)}) > a$ for all k), which contradicts (i).

(vi) Assume that $\mathfrak{C}(Z)$ is discrete. Then $\mathfrak{A}(x^{(n)})$ is discrete by (i), and we have from (iii) that $(x^{(n)})_n$ converges. Therefore, $\mathfrak{A}(x^{(n)})$ reduces to the singleton $\{\lim_{n \rightarrow \infty} x^{(n)}\}$, and hence, using (i) again, $\lim_{n \rightarrow \infty} x^{(n)} \in \mathfrak{C}(Z)$.

(vii) Assume that Z is strictly convex. Then, since Z is coercive, Z has a unique global minimum x^* and $\mathfrak{C}(Z) = \{x^*\}$. It follows from (vi) that $\lim_{n \rightarrow \infty} x^{(n)} = x^*$. \square

REFERENCES

- [1] E. Bae, J. Shi and X.-C. Tai, [Graph cuts for curvature based image denoising](#), *IEEE Trans. Image Process.*, **20** (2011), 1199–1210.
- [2] M. Banham and A. Katsaggelos, [Digital image restoration](#), *IEEE Signal Processing Mag.*, **14** (1997), 24–41.
- [3] L. Bedini, L. Benvenuti, E. Salerno and A. Tonazzini, [A mixed-annealing algorithm for edge preserving image reconstruction using a limited number of projections](#), *Signal Process.*, **32** (1993), 397–408.
- [4] M. Belge, M. Kilmer and E. Miller, [Wavelet domain image restoration with adaptive edge-preserving regularization](#), *IEEE Trans. Image Process.*, **9** (2000), 597–608.
- [5] M. Belge, M. Kilmer and E. Miller, [Efficient determination of multiple regularization parameters in a generalized L-curve framework](#), *Inverse Problems*, **18** (2002), 1161–1183.
- [6] L. Blanc-Féraud, S. Teboul, G. Aubert and M. Barlaud, [Nonlinear regularization using constrained edges in image reconstruction](#), in *Proc. IEEE Int. Conf. Image Processing*, **2**, Lausanne, Switzerland, (1996), 449–452.
- [7] J.-F. Cai, R. Chan and M. Nikolova, [Fast two-phase image deblurring under impulse noise](#), *J. Math. Imaging Vis.*, **36** (2010), 46–53.
- [8] R. Chan, Y. Dong and M. Hintermüller, [An efficient two-phase \$L^1\$ -TV method for restoring blurred images with impulse noise](#), *IEEE Trans. Image Process.*, **19** (2010), 1731–1739.
- [9] T. Chan, S. Kang and J. Shen, [Euler’s elastica and curvature-based inpainting](#), *SIAM J. Appl. Math.*, **63** (2002), 564–592.
- [10] T. Chan, A. Marquina and P. Mulet, [High-order total variation-based image restoration](#), *SIAM J. Sci. Comput.*, **22** (2000), 503–516.
- [11] P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud, [Deterministic edge-preserving regularization in computed imaging](#), *IEEE Trans. Image Process.*, **6** (1997), 298–311.
- [12] H. Choi and R. Baraniuk, [Wavelet statistical models and Besov spaces](#), in *Proc. SPIE, Wavelet Applications in Signal and Image Processing VII*, vol. 3813, 1999, 489–501.
- [13] E. Chouzenoux, J. Idier and S. Moussaoui, [A majorize-minimize strategy for subspace optimization applied to image restoration](#), *IEEE Trans. Image Process.*, **20** (2011), 1517–1528.
- [14] A. Cohen, I. Daubechies and J.-C. Feauveau, [Biorthogonal bases of compactly supported wavelets](#), *Comm. Pure Appl. Math.*, **45** (1992), 485–560.
- [15] I. Daubechies, M. Defrise and C. D. Mol, [An iterative thresholding algorithm for linear inverse problems with a sparsity constraint](#), *Comm. Pure Appl. Math.*, **57** (2004), 1413–1457.

- [16] A. H. Delaney and Y. Bresler, [Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography](#), *IEEE Trans. Image Process.*, **7** (1998), 204–221.
- [17] G. Demoment, [Image reconstruction and restoration: Overview of common estimation structures and problems](#), *IEEE Trans. Acoust. Speech Signal Process.*, **37** (1989), 2024–2036.
- [18] R. DeVore, [Nonlinear approximation](#), *Acta Numer.*, **7** (1998), 51–150.
- [19] D. Dobson and F. Santosa, [Recovery of blocky images from noisy and blurred data](#), *SIAM J. Appl. Math.*, **56** (1996), 1181–1198.
- [20] Y. Dong, M. Hintermüller and M. Neri, [An efficient primal-dual method for \$L^1\$ -TV image restoration](#), *SIAM J. Imaging Sci.*, **2** (2009), 1168–1189.
- [21] H. Farid and E. Simoncelli, [Differentiation of discrete multidimensional signals](#), *IEEE Trans. Image Process.*, **13** (2004), 496–508.
- [22] D. Geman and G. Reynolds, [Constrained restoration and the recovery of discontinuities](#), *IEEE Trans. Pattern Anal. Machine Intell.*, **14** (1992), 367–383.
- [23] S. Geman and D. Geman, [Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images](#), *IEEE Trans. Pattern Anal. Machine Intell.*, **6** (1984), 721–741.
- [24] S. Geman and D. McClure, [Bayesian image analysis: an application to single photon emission tomography](#), in *Proc. Stat. Comput. Section: annual meeting of the Amer. Stat. Assoc.*, Las Vegas, Nevada, 1985, 12–18.
- [25] P. Green, [Bayesian reconstructions from emission tomography data using a modified EM algorithm](#), *IEEE Trans. Med. Imag.*, **9** (1990), 84–93.
- [26] T. Hebert and R. Leahy, [A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors](#), *IEEE Trans. Med. Imag.*, **8** (1989), 194–202.
- [27] T. Hou, S. Wang and H. Qin, [Image deconvolution with multi-stage convex relaxation and its perceptual evaluation](#), *IEEE Trans. Image Process.*, **20** (2011), 1057–1149.
- [28] Y. Hu and M. Jacob, [Higher degree total variation \(HDTV\) regularization for image recovery](#), *IEEE Trans. Image Process.*, **21** (2012), 2559–2571.
- [29] J. Idier, [Convex half-quadratic criteria and interacting auxiliary variables for image restoration](#), *IEEE Trans. Image Process.*, **10** (2001), 1001–1009.
- [30] K. Ito, B. Jin and T. Takeuchi, [A regularization parameter for nonsmooth Tikhonov regularization](#), *SIAM J. Sci. Comput.*, **33** (2011), 1415–1438.
- [31] P. Kohli, A. Shekhovtsov, C. Rother, V. Kolmogorov and P. Torr, [On partial optimality in multi-label MRFs](#), in *Proc. 25th Int. Conf. Machine Learning*, Helsinki, Finland, 2008, 480–487.
- [32] S. Lefkimmiatis, A. Bourquard and M. Unser, [Hessian-based norm regularization for image restoration with biomedical applications](#), *IEEE Trans. Image Process.*, **21** (2012), 983–995.
- [33] S. Li, [On discontinuity-adaptive smoothness priors in computer vision](#), *IEEE Trans. Pattern Anal. Machine Intell.*, **17** (1995), 576–586.
- [34] Y.-R. Li, L. Shen, D.-Q. Dai and B. Suter, [Framelet algorithms for de-blurring images corrupted by impulse plus Gaussian noise](#), *IEEE Trans. Image Process.*, **20** (2011), 1822–1837.
- [35] M. Lysaker and X.-C. Tai, [Iterative image restoration combining total variation minimization and a second-order functional](#), *Int. J. Comput. Vis.*, **66** (2006), 5–18.
- [36] J. Ma and G. Plonka, [The curvelet transform](#), *IEEE Signal Processing Mag.*, **27** (2010), 118–133.
- [37] S. Mallat, [A theory for multiresolution signal decomposition: the wavelet representation](#), *IEEE Trans. Pattern Anal. Machine Intell.*, **11** (1989), 674–693.
- [38] J. Marroquin, S. Mitter and T. Poggio, [Probabilistic solution of ill-posed problems in computational vision](#), *J. Amer. Statist. Assoc.*, **82** (1987), 76–89.
- [39] R. Meyer, [Sufficient conditions for the convergence of monotonic mathematical programming algorithms](#), *J. Comput. System Sci.*, **12** (1976), 108–121.
- [40] P. Moulin and J. Liu, [Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors](#), *IEEE Trans. Inform. Theory*, **45** (1999), 909–919.
- [41] M. Nikolova, [A variational approach to remove outliers and impulse noise](#), *J. Math. Imaging Vis.*, **20** (2004), 99–120.
- [42] M. Nikolova, [Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares](#), *Multiscale Model. Simul.*, **4** (2005), 960–991.
- [43] M. Nikolova, M. Ng and C.-P. Tam, [Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction](#), *IEEE Trans. Image Process.*, **19** (2010), 3073–3088.

- [44] S. Ramani, T. Blu and M. Unser, [Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms](#), *IEEE Trans. Image Process.*, **17** (2008), 1540–1554.
- [45] M. C. Robini and I. Magnin, [Optimization by stochastic continuation](#), *SIAM J. Imaging Sci.*, **3** (2010), 1096–1121.
- [46] M. C. Robini, T. Rastello and I. Magnin, [Simulated annealing, acceleration techniques and image restoration](#), *IEEE Trans. Image Process.*, **8** (1999), 1374–1387.
- [47] P. Rodríguez and B. Wohlberg, [Efficient minimization method for a generalized total variation functional](#), *IEEE Trans. Image Process.*, **18** (2009), 322–332.
- [48] C. Rother, V. Kolmogorov, V. Lempitsky and M. Szummer, [Optimizing binary MRFs via extended roof duality](#), in *Proc. IEEE Conf. Comp. Vis. Patt. Recognition*, Minneapolis, MN, 2007, 1–8.
- [49] X.-C. Tai, J. Hahn and G. Chung, [A fast algorithm for Euler’s elastica model using augmented Lagrangian method](#), *SIAM J. Imaging Sci.*, **4** (2011), 313–344.
- [50] M. Unser, [Ten good reasons for using spline wavelets](#), in *Proc. SPIE, Wavelet Applications in Signal and Image Processing V*, vol. 3169, 1997, 422–431.
- [51] L. Wang, T.-T. Wong, P. Heng and J. Cheng, [Template-matching approach to edge detection of volume data](#), in *Proc. Int. Workshop on Medical Imaging and Augmented Reality*, Shatin, Hong Kong, 2001, 286–291.
- [52] T. Zhang, [Analysis of multi-stage convex relaxation for sparse regularization](#), *J. Mach. Learn. Res.*, **11** (2010), 1081–1107.
- [53] W. Zhu and T. Chan, [Image denoising using mean curvature of image surface](#), *SIAM J. Imaging Sci.*, **5** (2012), 1–32.
- [54] X. Zhu and P. Milanfar, [Automatic parameter selection for denoising algorithms using a no-reference measure of image content](#), *IEEE Trans. Image Process.*, **19** (2010), 3116–3132.

Received March 2012; revised March 2013.

E-mail address: marc.robini@creatis.insa-lyon.fr

E-mail address: zhu@creatis.insa-lyon.fr

E-mail address: jhluo@sjtu.edu.cn