Image Restoration Subject to a Total Variation Constraint

Patrick L. Combettes, *Senior Member, IEEE,* and Jean-Christophe Pesquet, *Senior Member, IEEE*

***Abstract*—Total variation has proven to be a valuable concept in connection with the recovery of images featuring piecewise smooth components. So far, however, it has been used exclusively as an objective to be minimized under constraints. In this paper, we propose an alternative formulation in which total variation is used as a constraint in a general convex programming framework. This approach places no limitation on the incorporation of additional constraints in the restoration process and the resulting optimization problem can be solved efficiently via block-iterative methods. Image denoising and deconvolution applications are demonstrated.**

I. Problem Statement

THE CLASSICAL linear restoration problem is to find the original form of an image $\bar{x}$ in a real Hilbert space $(\mathcal{H}, \|\cdot\|)$ from the observation of a degraded image

$$y = L\bar{x} + u \tag{1}$$

where $L\colon \mathcal{H} \to \mathcal{H}$ is a bounded linear operator modeling the blurring process and $u \in \mathcal{H}$ models an additive noise component. Numerous approaches have been developed over the past three decades to solve this problem (see, for instance, [2], [9], [16], [33], [35], [39], and the references therein). Roughly speaking, restoration problems are typically posed as optimization problems in which an appropriate objective function is minimized under certain constraints. Restricting ourselves to convex problems, a general formulation is, therefore

$$\text{Find } x \in S = \bigcap_{i=1}^{m} S_i \quad \text{such that } J(x) = \inf J(S) \tag{2}$$

where the objective $J : \mathcal{H} \to]-\infty, +\infty]$ is a convex function and the constraint sets $(S_i)_{1\le i\le m}$ are closed convex subsets of $\mathcal{H}$. These constraints arise from *a priori* knowledge about the model (1) and the original image $\bar{x}$. For instance, the classical formulation of [21] concerns problems with smooth images in which the energy δ of the noise u is known. The goal is then to find the smoothest image in terms of some high-pass filtering operator $C : \mathcal{H} \to \mathcal{H}$ which is consistent with (1) and the noise information, hence $J : x \mapsto \|Cx\|^2$ and $S = \{x \in \mathcal{H} \mid \|Lx - y\|^2 \le \delta\}$ in (2). In the Sobolev space $\mathcal{H} = W^{2,2}(\Omega)$, where Ω is an open subset of $\mathbb{R}^2$, the formulation of [21] reads

$$\text{minimize} \int_\Omega |\Delta x(\omega)|^2 d\omega \text{ subject to } \|Lx - y\|^2 \le \delta \tag{3}$$

where Δx denotes the Laplacian of x. In $\mathcal{H} = W^{1,2}(\Omega)$, smoothness can also be enforced with the objective $J : x \mapsto \int_\Omega |\nabla x(\omega)|_2^2 d\omega$, where $|\cdot|_2$ denotes the Euclidean norm in $\mathbb{R}^2$ and ∇x the gradient of x [20], [22], [34].

A. Total Variation Restoration

For images with sharp contours and block features, it was proposed in [32] (see also [26] and [31]) to formulate the restoration problem as

$$\text{minimize} \int_\Omega |\nabla x(\omega)|_2 d\omega \text{ subject to } \|Lx - y\|^2 \le \delta. \tag{4}$$

If $x \in W^{1,1}(\Omega)$, the quantity

$$\mathrm{tv}(x) = \int_\Omega |\nabla x(\omega)|_2 d\omega \tag{5}$$

is the total variation of x; physically, $\mathrm{tv}(x)$ is a measure of the amount of oscillations in x. It should be noted that, under certain assumptions [7], Lagrangian theory provides a conceptual equivalence between (4) and the unconstrained problem

$$\text{minimize } \mathrm{tv}(x) + \lambda\|Lx - y\|^2, \quad \lambda \ge 0 \tag{6}$$

which is often implemented in practice (see also [30] for alternative Lagrangian formulations). However, finding the exact Lagrange parameter λ is a computationally intensive task. As a result, the numerical value of λ in (6) is typically set in an *ad hoc* manner, e.g., [23], [38], which can have a significant impact on the solution [1].

The formulation (4), and its variant (6), has proven particularly effective in denoising problems for piecewise smooth images, and it has been the focus of a great deal of attention in recent years in the image processing and applied mathematics communities, e.g., [1], [5]–[8], [12], [14], [19], [23], [24], [28], [30], and [38]. At the same time, it has some inherent limitations.

- **Staircase effect:** The intensity levels of images produced by total variation minimization tend to cluster in patches. These artifacts have been observed by several authors, e.g., [8], [39], and investigated for one-dimensional discrete models in [25]. An explanation for the staircase effect is that the algorithm produces a restored image x whose total

Manuscript received June 18, 2003; revised December 3, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Giovanni Ramponi.

P. L. Combettes is with the Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie - Paris 6, 75005 Paris, France (e-mail: plc@math.jussieu.fr).

J.-C. Pesquet is with the Institut Gaspard Monge and Centre National de la Recherche Scientifique (UMR–CNRS 8049), Université de Marne-la-Vallée, 77454 Marne-la-Vallée, France (e-mail: pesquet@univ-mlv.fr).

Digital Object Identifier 10.1109/TIP.2004.832922

variation can be significantly below that of the original image $\bar{x}$.

- **Knowledge of noise environment:** The construction of the bound δ in (4) requires specific assumptions and information about the noise [15], [36]. In some problems, such conditions may not be met.

In connection with the question of solving (4) exactly via numerical algorithms, it should also be noted that the nondifferentiable optimization algorithms currently in use are not easily amenable to the incorporation of additional complex constraints [14].

B. Proposed Approach

The goal of this paper is to propose an alternative framework for incorporating total variation in the restoration process. Our starting point is the observation that, in the spirit of feasibility methods [9], [36], [40], one would ideally like to obtain a restored image in the convex set $\{x \in \mathcal{H} \,|\, \mathrm{tv}(x) \leq \mathrm{tv}(\bar{x})\}$ of images whose total variation is consistent with that of the original image $\bar{x}$. Of course, in practice, $\mathrm{tv}(\bar{x})$ is usually unknown, but a reasonable upper bound τ may be available for certain classes of images based on statistics of databases. Thus, under suitable assumptions, if $\bar{x}$ is a binary image, then $\mathrm{tv}(\bar{x})$ is the length of the boundary of its support [18, example 1.4]; if $\bar{x}$ is a gray-level image, then $\mathrm{tv}(\bar{x})$ is obtained by integrating the length of the level curves $\{\omega \in \Omega \,|\, \bar{x}(\omega) = s\}$ with respect to s [41, theorem 2.7.1] (see also [41, theorem 5.4.4]). Hence, $\mathrm{tv}(\bar{x})$ constitutes a geometrical attribute that can be expected to exhibit limited variance over certain classes of images, e.g., views of similar urban areas in satellite imaging, tomographic reconstructions of similar cross sections in tomography, fingerprint images, text images, or face images. For instance, on a database of 564 images showing different (from 19 to 48) views of 20 different human faces [37], we have computed the average value $\overline{\mathrm{tv}}$ of the total variation for each person as well as the proportion of views whose total variation lies between 0.8 $\overline{\mathrm{tv}}$ and 1.2 $\overline{\mathrm{tv}}$ (as will be seen in Section IV-A.3, this typically corresponds to the kind of deviation which is acceptable in our framework). Fig. 1 shows that the proportion of images between these bounds is around 99%.

The information $\mathrm{tv}(\bar{x}) \leq \tau$ restricts the solutions to the convex set

$$S_1 = \{x \in \mathcal{H} \,|\, \mathrm{tv}(x) \leq \tau\}. \tag{7}$$

Given a pertinent objective J, this set can be incorporated in (2) in conjunction with additional constraints $(S_i)_{2 \leq i \leq m}$ arising from *a priori* knowledge. This framework departs from the traditional formulation (4) in that total variation is now used as a constraint rather than an objective. Under mild assumptions on J, the resulting optimization problem (2) is much more attractive since, as a constraint, total variation can be efficiently processed via subgradient projections in the context of block-iterative outer approximation methods [11]. Furthermore, this approach places no restriction on the number of constraints m or on their analytical complexity.

It should be noted that the idea of using an upper bound on the total variation of images also appears in [27] as a stopping criterion in a PDE-driven restoration scheme.

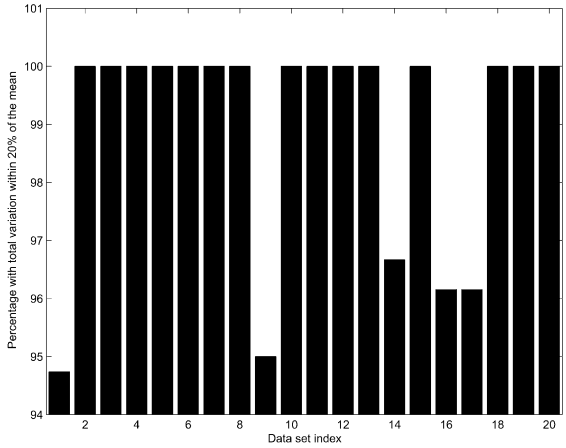


Fig. 1. For 20 different data sets, the percentage of images whose total variation falls within 20% of the average total variation of the set.

C. Organization of the Paper

In Section II, we provide the necessary background material. In Section III, the proposed problem formulation is formally stated and a block-iterative parallel algorithm is described to solve it. Two numerical applications are considered in Section IV, namely deconvolution of satellite images and denoising in the presence of non-Gaussian noise. The proofs of technical results are placed in the Appendix.

II. Background

A. Notation

The underlying image space is a real Hilbert space $\mathcal{H}$ with scalar product $\langle \cdot \,|\, \cdot \rangle$ and norm $\| \cdot \|$. The distance from an image $x \in \mathcal{H}$ to a nonempty set $C \subset \mathcal{H}$ is $d_C(x) = \inf \|x - C\|$. Given a continuous convex function $f : \mathcal{H} \to \mathbb{R}$ and $\eta \in \mathbb{R}$, the closed and convex set

$$\mathrm{lev}_{\leq \eta} f = \{x \in \mathcal{H} \,|\, f(x) \leq \eta\} \tag{8}$$

is the lower level set of f at height η. A vector $t \in \mathcal{H}$ is a subgradient of f at $x \in \mathcal{H}$ if

$$(\forall z \in \mathcal{H})\ \langle z - x \,|\, t \rangle + f(x) \leq f(z). \tag{9}$$

As f is continuous, it always possesses at least one subgradient at each point $x \in \mathcal{H}$. If f is (Gâteaux) differentiable at x, then it possesses a unique subgradient at this point, namely its gradient $\nabla f(x)$. The set of all subgradients of f at x is the subdifferential of f at x and is denoted by $\partial f(x)$. For background on convex analysis, see [17] and [29].

B. Subgradient Projections

The reader will find a more detailed account of subgradient projections in [4], [9], and [10]; we provide only the essential facts here.

Let C be a nonempty closed and convex subset of $\mathcal{H}$ and let x be a point in $\mathcal{H}$. Then, there exists a unique point $P_C x \in C$ such that $\|x - P_C x\| = d_C(x)$; $P_C x$ is called the projection of x onto C. Now, suppose that $C = \mathrm{lev}_{\leq 0} f \neq \emptyset$, where $f : \mathcal{H} \to \mathbb{R}$ is continuous and convex (the identity $C = \mathrm{lev}_{\leq 0} d_C$ shows that

such a representation always exists). Fix $x \in \mathcal{H}$, a subgradient $t \in \partial f(x)$, and define

$$H_x = \begin{cases} \{z \in \mathcal{H} \,|\, \langle x - z \,|\, t\rangle \geq f(x)\}, & \text{if } f(x) > 0 \\ \mathcal{H}, & \text{if } f(x) \leq 0. \end{cases} \quad (10)$$

Then, $C \subset H_x$ and the projection of x onto H_x, i.e.

$$G_C x = P_{H_x} x = \begin{cases} x - \dfrac{f(x)t}{\|t\|^2}, & \text{if } f(x) > 0 \\ x, & \text{if } f(x) \leq 0 \end{cases} \quad (11)$$

is called a subgradient projection of x onto C. We note that computing $G_C x$ requires only the knowledge of a subgradient of f at x and is, therefore, much more economical than computing the exact projection $P_C x$, as the latter amounts to solving a constrained quadratic minimization problem. However, when $P_C x$ is easy to compute, one can set $f = d_C$ and obtain $G_C x = P_C x$.

C. Image Spaces and Total Variation

Complements on functional spaces and total variation will be found in [18] and [41].

Let Ω be an open subset of $\mathbb{R}^2$, let $p \in [1, +\infty[$, and let $\omega = (\omega_1, \omega_2)$ be a generic point in Ω. As usual, $\mathcal{D}(\Omega)$ denotes the space of infinitely differentiable functions with compact support in Ω, and $L^p(\Omega)$ is the space of functions whose pth power is absolutely integrable on Ω. Given $x \in L^p(\Omega)$ and $i \in \{1,2\}$, the function $D_i x : \Omega \to \mathbb{R}$ is the weak partial derivative of x with respect to ω_i if $(\forall \varphi \in \mathcal{D}(\Omega))\ \int_\Omega \varphi(\omega) D_i x(\omega) d\omega = -\int_\Omega (\partial\varphi(\omega)/\partial\omega_i) x(\omega) d\omega$; the (weak) gradient of x is $\nabla x = (D_1 x, D_2 x)$. The Sobolev space of index $(1,p)$ is $W^{1,p}(\Omega) = \{x \in L^p(\Omega) \,|\, D_i x \in L^p(\Omega), i = 1,2\}$; $H^1(\Omega) = W^{1,2}(\Omega)$ is a Hilbert space. Likewise, $W^{2,p}(\Omega) = \{x \in W^{1,p}(\Omega) \,|\, D_i x \in W^{1,p}(\Omega), i = 1,2\}$.

It will, henceforth, be assumed that $\Omega \subset \mathbb{R}^2$ is a nonempty bounded open set. An analog image is modeled as an element in $H^1(\Omega)$. Since Ω is bounded, $H^1(\Omega) \subset W^{1,1}(\Omega)$ and the total variation of an image $x \in H^1(\Omega)$ is finite and given by (5).

On the other hand, in the discrete setting that prevails in numerical applications, analog images are discretized on a finite $M \times M$ sampling grid. The total variation of the discrete image $x = [x^{i,j}]_{0 \leq i,j \leq M-1} \in \mathbb{R}^{M \times M}$ is obtained by discretizing (5) into (see [14])

$$\begin{aligned} \mathrm{tv}(x) = &\sum_{i,j=0}^{M-2} \sqrt{|x^{i+1,j} - x^{i,j}|^2 + |x^{i,j+1} - x^{i,j}|^2} \\ &+ \sum_{i=0}^{M-2} |x^{i+1,M-1} - x^{i,M-1}| \\ &+ \sum_{j=0}^{M-2} |x^{M-1,j+1} - x^{M-1,j}|. \end{aligned} \quad (12)$$

Now, set $N = M^2$. As is customary, a discrete image $x = [x^{i,j}]_{0 \leq i,j \leq M-1} \in \mathbb{R}^{M \times M}$ will be dealt with as a vector in the usual Euclidean space $\mathcal{H} = \mathbb{R}^N$ through the column stacking isometry $x^{i,j} \leftrightarrow x^{i+Mj}$ [2]. In turn, upon introducing suitable difference matrices $(\Gamma_{i,j})_{0 \leq i,j \leq M-2}$ in $\mathbb{R}^{2 \times N}$, $(\Gamma_{i,M-1})_{0 \leq i \leq M-2}$ in $\mathbb{R}^{1 \times N}$, and $(\Gamma_{M-1,j})_{0 \leq j \leq M-2}$ in $\mathbb{R}^{1 \times N}$, the total variation of $x \in \mathbb{R}^N$ can be expressed as

$$\mathrm{tv}(x) = \sum_{i,j=0}^{M-2} |\Gamma_{i,j} x|_2 + \sum_{i=0}^{M-2} |\Gamma_{i,M-1} x| + \sum_{j=0}^{M-2} |\Gamma_{M-1,j} x|. \quad (13)$$

Proposition 1: Let $\mathcal{H}$ be either $H^1(\Omega)$ or $\mathbb{R}^N$. Then, $\mathrm{tv} : \mathcal{H} \to \mathbb{R}$ is a continuous convex function.

III. Numerical Method

A. Assumptions

As formulated in Section I-B, the restoration problem is to solve (2) in a Hilbert space $\mathcal{H}$, where $J : \mathcal{H} \to]-\infty, +\infty]$ is a convex function, S_1 is given by (7), and $(S_i)_{2 \leq i \leq m}$ are closed convex sets in $\mathcal{H}$. We saw in Section II-C that $\mathcal{H}$ is either $H^1(\Omega)$ or $\mathbb{R}^N$, according as we consider the theoretical analog model or the discrete model. As in [10] and [13], it is convenient to model the constraint sets $(S_i)_{2 \leq i \leq m}$ as level sets, say

$$(\forall i \in \{2, \ldots, m\})\ S_i = \mathrm{lev}_{\leq 0} f_i \quad (14)$$

where $(f_i)_{2 \leq i \leq m}$ are continuous convex functions from $\mathcal{H}$ to $\mathbb{R}$. In view of Proposition 1, the set S_1 can also be put in this format by setting $f_1 = \mathrm{tv} - \tau$. The problem is, therefore, to

$$\text{minimize } J(x) \text{ subject to } \max_{1 \leq i \leq m} f_i(x) \leq 0. \quad (15)$$

We shall use the general convex programming framework developed in [11], which is particularly well adapted to problems of the type specified by (2) and (14). In order to avoid technical complications, we shall focus on the numerical results of immediate computational interest in $\mathcal{H} = \mathbb{R}^N$ (the reader can refer to [11] for the infinite dimensional results pertinent to the convergence analysis in $H^1(\Omega)$). We now state our assumptions formally.

Assumption 2:

1) $\mathcal{H} = \mathbb{R}^N$, S_1 is defined by (7), $(S_i)_{2 \leq i \leq m}$ by (14), where the functions $(f_i)_{2 \leq i \leq m}$ are finite and convex, and $S = \bigcap_{i=1}^m S_i \neq \emptyset$.
2) $J : \mathbb{R}^N \to]-\infty, +\infty]$ is convex and lower semicontinuous.
3) There exists $z \in S$ such that $J(z) < +\infty$, $C = \mathrm{lev}_{\leq J(z)} J$ is bounded and J is differentiable and strictly convex on C.

Proposition 3: Suppose that $S \neq \emptyset$ and that $J : \mathbb{R}^N \to \mathbb{R}$ is strictly convex, differentiable, and coercive, that is, $\lim_{\|x\| \to +\infty} J(x) = +\infty$. Then items 2)–3) in Assumption 2 are satisfied.

B. Solution Algorithm

Recall from (11) that, if $t_{i,n}$ is a subgradient of f_i at $x_n \in \mathbb{R}^N$, the subgradient projection of x_n onto S_i associated with $t_{i,n}$ is

$$p_{i,n} = \begin{cases} x_n - \dfrac{f_i(x_n) t_{i,n}}{\|t_{i,n}\|^2}, & \text{if } f_i(x_n) > 0 \\ x_n, & \text{if } f_i(x_n) \leq 0. \end{cases} \quad (16)$$

Given $I_n \subset \{1,\dots,m\}$ and convex weights $(\omega_{i,n})_{i\in I_n}$, define

$$\lambda_n = \begin{cases} \dfrac{\sum_{i\in I_n}\omega_{i,n}\|p_{i,n}-x_n\|^2}{\|x_n-\sum_{i\in I_n}\omega_{i,n}p_{i,n}\|^2}, & \text{if } \max\limits_{i\in I_n} f_i(x_n) > 0 \\ 1, & \text{otherwise.} \end{cases} \tag{17}$$

We are now ready to describe the algorithm.

Algorithm 4
① Fix $\varepsilon \in]\,0, 1/m\,[$. Let x_0 be the minimizer of J over $\mathbb{R}^N$ and set $n = 0$.
② Take a nonempty index set $I_n \subset \{1,\dots,m\}$.
③ Set $z_n = x_n + \lambda_n(\sum_{i\in I_n}\omega_{i,n}p_{i,n} - x_n)$, where
 a) for every $i \in I_n$, $p_{i,n}$ is as in (16);
 b) $(\omega_{i,n})_{i\in I_n}$ lies in $]\varepsilon, 1]$ and $\sum_{i\in I_n}\omega_{i,n} = 1$;
 c) λ_n is as in (17).
④ Set $D_n = \{x \in \mathbb{R}^N \,|\, \langle x_n - x \,|\, \nabla J(x_n)\rangle \le 0\}$ and $H_n = \{x \in \mathbb{R}^N \,|\, \langle x - z_n \,|\, x_n - z_n\rangle \le 0\}$.
⑤ Let x_{n+1} be the minimizer of J over $D_n \cap H_n$.
⑥ Set $n = n+1$ and go to ②.

Theorem 5: Suppose that Assumption 2 is satisfied and that there exists a strictly positive integer M such that, for every $i \in \{1,\dots,m\}$ and every $n \in \mathbb{N}$, $i \in \bigcup_{k=n}^{n+M-1} I_k$. Then, every sequence generated by Algorithm 4 converges to the unique solution to (2) or, equivalently, to (15).

The key step in the algorithm is Step ⑤, in which the update x_{n+1} is obtained as the minimizer of J over the intersection of the two half spaces D_n and H_n, which contains S. Specific implementations of this basic minimization problem under two affine inequality constraints will be discussed in Sections IV-A.2 and IV-B.2. Let us note that if one bypasses Step ④ and replaces Step ⑤ by $x_{n+1} = z_n$, one recovers an algorithm proposed in [10]. However, in this case, $(x_n)_{n\ge 0}$ converges to an unspecified image in the feasibility set S [10, theorem 3].

C. Practical Considerations

A few remarks are in order.

1) If the total variation set S_1 of (7) is selected at iteration n, the subgradient projection $p_{1,n}$ of x_n is required. In view of (16) and (13), we have

$$p_{1,n} = \begin{cases} x_n - \dfrac{(\mathrm{tv}(x_n)-\tau)t_{1,n}}{\|t_{1,n}\|^2}, & \text{if } \mathrm{tv}(x_n) > \tau \\ x_n, & \text{if } \mathrm{tv}(x_n) \le \tau \end{cases} \tag{18}$$

where the subgradient $t_{1,n} \in \partial\,\mathrm{tv}(x_n)$ can be taken to be (see Section IV-A in [14])

$$t_{1,n} = \sum_{i,j=0}^{M-2} u_{i,j} + \sum_{i=0}^{M-2} v_i + \sum_{j=0}^{M-2} w_j \tag{19}$$

with (20), as shown at the bottom of the page.

2) The condition on $(I_n)_{n\ge 0}$ imposes merely that every index i be used at least once within any M consecutive iterations. As a result, Algorithm 4 is very flexible in terms of parallel implementation. For instance, if m parallel processors are available, one can choose to use all the sets simultaneously at iteration n, i.e., $I_n = \{1,\dots,m\}$; if only one processor is available, one can sweep through the sets periodically as in the POCS algorithm [40], i.e., $I_n = \{(n \text{ modulo } m) + 1\}$. Intermediate block-iterative sweeping rules are described in [10].
3) If the projection $P_{S_i}x_n$ of x_n onto a selected set S_i is easy to compute, one can take $p_{i,n} = P_{S_i}x_n$, which amounts to setting $f_i = d_{S_i}$.

Fig. 2. Original image. Pixel range [0, 255].

IV. Simulations Results

A. Application to Satellite Imaging

1) Experiments: The original image $\bar{x}$ shown in Fig. 2 is a SPOT-5 256×256 satellite image, hence $\mathcal{H} = \mathbb{R}^N$, where $N = 256^2$. The degraded image y shown in Fig. 3 is obtained

$$\begin{cases} u_{ij} = \begin{cases} 0, & \text{if } \Gamma_{i,j}x_n = 0 \\ \Gamma_{i,j}^\top \Gamma_{i,j}x_n/|\Gamma_{i,j}x_n|_2, & \text{otherwise} \end{cases} \\ v_i = \begin{cases} 0, & \text{if } \Gamma_{i,M-1}x_n = 0 \\ \Gamma_{i,M-1}^\top \Gamma_{i,M-1}x_n/|\Gamma_{i,M-1}x_n|, & \text{otherwise} \end{cases} \\ w_j = \begin{cases} 0, & \text{if } \Gamma_{M-1,j}x_n = 0 \\ \Gamma_{M-1,j}^\top \Gamma_{M-1,j}x_n/|\Gamma_{M-1,j}x_n|, & \text{otherwise.} \end{cases} \end{cases} \tag{20}$$

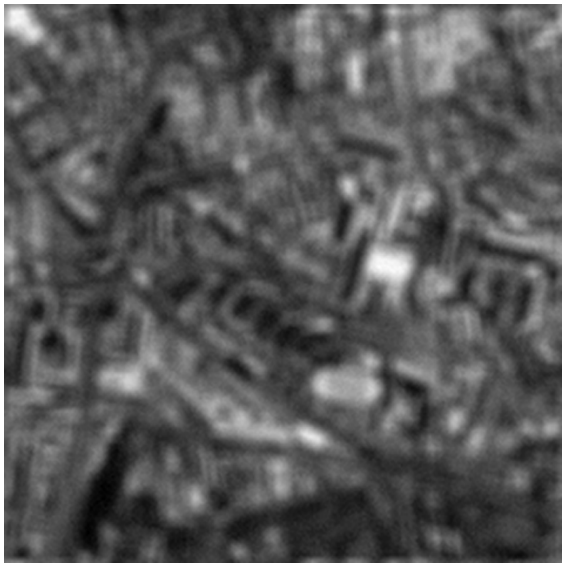


Fig. 3. Degraded image. Pixel range [4, 263].

Fig. 5. Quadratic deconvolution with bounded total variation. Pixel range [−55, 316].

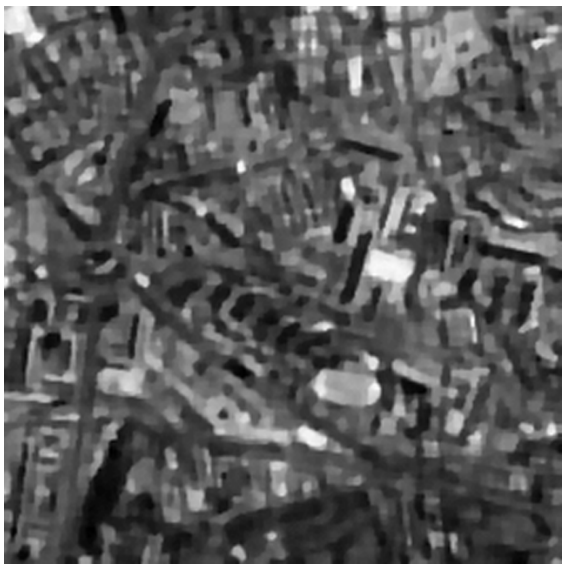


Fig. 4. Standard total variation deconvolution by (4). Pixel range [0, 267].

Fig. 6. Quadratic deconvolution with bounded total variation and additional constraints. Pixel range [0, 259].

by convolving $\bar{x}$ with a 7×7 uniform blur and adding Gaussian white noise. The blurred-image to noise ratio is 30 dB. The degradation model is, therefore, (1), where the convolution matrix L is assumed to be known.

First, we assume that the above mentioned characteristics of the noise are known. The bound δ of (4) can, therefore, be derived from these informations [15], [36]. The standard total variation problem (4) is solved using the adaptive level set method of [14] and its solution is shown in Fig. 4. The total variation of the restored image is 9.151×10^5, which is much below the true value $\mathrm{tv}(\bar{x}) = 1.774 \times 10^6$.

We now turn to an alternative scenario, in which no information is available about the noise but a bound τ is available on the total variation of the original image $\bar{x}$ (e.g., as estimated from a pool of similar satellite images), which makes it possible to set up the problem in the proposed format (2). The objective is chosen to be $J : x \mapsto \|Lx - y\|^2 + \alpha\|x\|^2$, where $\alpha = 10^{-3}$. Note that since L is not invertible, the residual energy function $x \mapsto \|Lx - y\|^2$ is not strictly convex. We, therefore, introduce the term $\alpha\|x\|^2$ to make J strictly convex, in compliance with item 3) in Assumption 2. Assuming no further information on x leads to the minimization of J over S_1. The solution produced by Algorithm 4 in this case has total variation 1.775×10^6 and is shown in Fig. 5. Next, we assume that additional information is available about $\bar{x}$, namely amplitude bounds and mean. Hence, we use the set S_1 of (7), as well as the sets $S_2 = [0, 255]^N$ and $S_3 = \{x \in \mathbb{R}^N \mid \langle x \mid \underline{1} \rangle = \mu\}$ in (2), where $\underline{1}$ is the vector whose entries are all equal to 1. The minimizer of J over $\bigcap_{i=1}^{3} S_i$, obtained by Algorithm 4, has total variation 1.777×10^6 and is shown in Fig. 6. It is apparent that the incorporation of additional information has improved the restoration. Finally, to provide evidence of the value of total variation information, we display in Fig. 7 the image obtained by minimizing J over $S_2 \cap S_3$, i.e.,

Fig. 7. Quadratic deconvolution with no total variation constraint. Pixel range [0, 257].

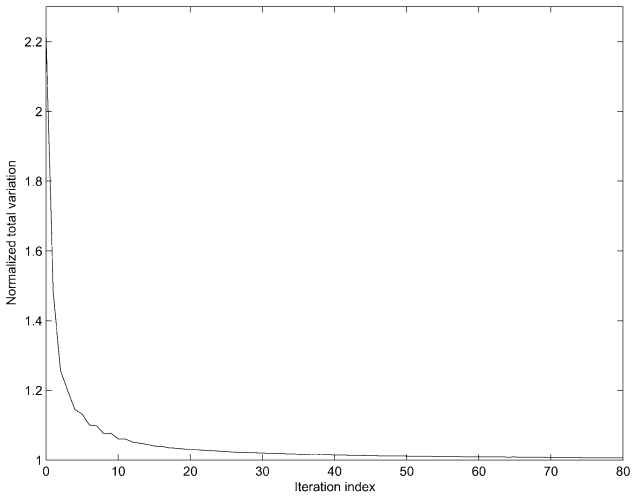


Fig. 8. $\mathrm{tv}(x_n)/\mathrm{tv}(\bar{x})$ versus the iteration index n.

without the total variation constraint. This image has total variation 3.8231×10^6.

2) Implementation of Algorithm 4 and Convergence Patterns: Exact projections onto S_2 and S_3 are used since they are easily computed, whereas a subgradient projection onto S_1 is employed. Now, let $R = L^\top L + \alpha \mathrm{Id}$. Then, the initial point is seen to be $x_0 = R^{-1}L^\top y$. Moreover, since the objective J is quadratic, the expression of x_{n+1} at Step ⑤ of Algorithm 4 can be obtained in closed form [13]. Thus, the nth iteration can be executed as described in the following pseudocode, where the subscript n has been dropped (see Section IV-C in [13] for details).

1) For every $i \in I$, set $a_i = -f_i(x)t_i/\|t_i\|^2$, where $t_i \in \partial f_i(x)$, if $f_i(x) > 0$; $a_i = 0$ otherwise.
2) Choose convex weights $(\omega_i)_{i\in I}$ conforming to Step ③b). Set $v = \sum_{i\in I}\omega_i a_i$ and $\lambda = \sum_{i\in I}\omega_i\|a_i\|^2$.
3) If $\lambda = 0$, exit iteration. Otherwise, set $b = x_0 - x$, $c = Rb$, $d = R^{-1}v$, and $\lambda = \lambda/\langle d \,|\, v\rangle$.

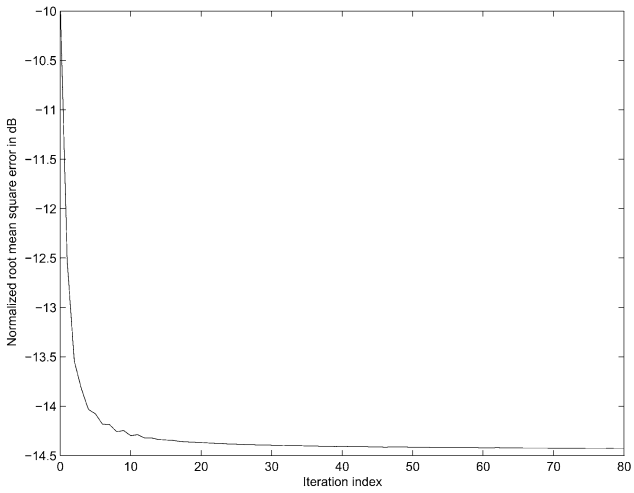


Fig. 9. $\|x_n - \bar{x}\|/\|\bar{x}\|$ versus the iteration index n.

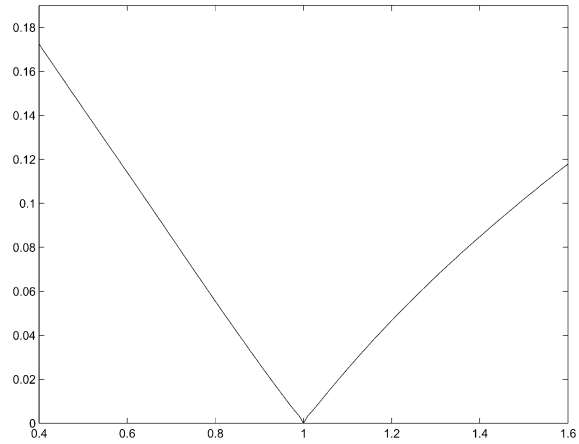


Fig. 10. Normalized error $\|x_\tau - x_{\bar{\tau}}\|/\|x_{\bar{\tau}}\|$ as a function of $\tau/\bar{\tau}$.

4) Set $d = \lambda d$.
5) Set $\pi = -\langle c \,|\, d\rangle$, $\mu = \langle b \,|\, c\rangle$, $\nu = \lambda\langle d \,|\, v\rangle$, and $\rho = \mu\nu - \pi^2$.
6) Set

$$x = \begin{cases} x + d, & \text{if } \rho = 0 \text{ and } \pi \geq 0 \\ x_0 + (1 + \pi/\nu)d, & \text{if } \rho > 0 \text{ and } \pi\nu \geq \rho \\ x + \dfrac{\nu}{\rho}(\pi b + \mu d), & \text{if } \rho > 0 \text{ and } \pi\nu < \rho. \end{cases}$$

To illustrate the convergence behavior of the method in the case of the last experiment corresponding to Fig. 6, we display the evolution of the total variation and of the root mean-square error in Figs. 8 and 9. Both indicate that convergence occurs roughly around the 50th iteration.

3) Sensitivity Assessment: In practice, the value of τ may not be known precisely and it is, therefore, important to evaluate its impact on the solution. Let us denote by x_τ the solution produced by the algorithm with the constraint sets $(S_i)_{1\leq i\leq 3}$ for a given bound τ in (7) and by $\bar{\tau} = \mathrm{tv}(\bar{x})$ the true bound. The normalized root mean square error $\|x_\tau - x_{\bar{\tau}}\|/\|x_{\bar{\tau}}\|$ as a function of $\tau/\bar{\tau}$ is displayed in Fig. 10. This curve shows that the method is robust in the sense that, as $\tau/\bar{\tau}$ varies from 0.82 to 1.21, the resulting error does not exceed 5%.

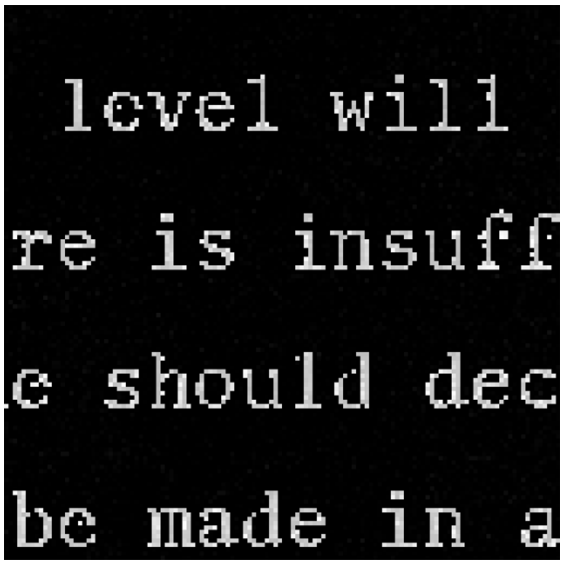


Fig. 11. Original image. Pixel range [0, 255].

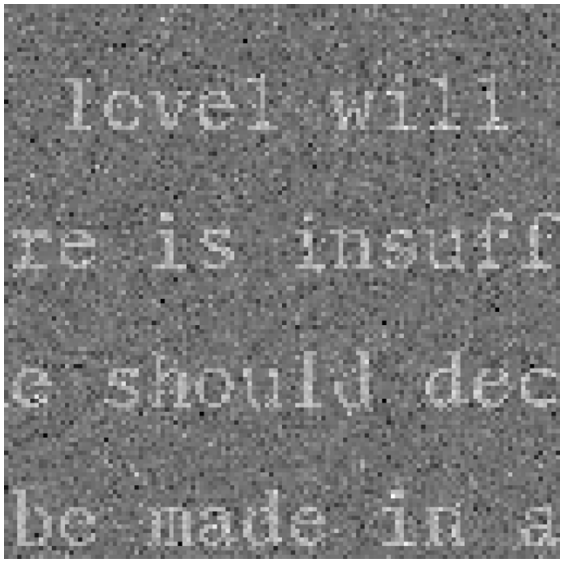


Fig. 12. Noisy image. Pixel range [−261, 460].

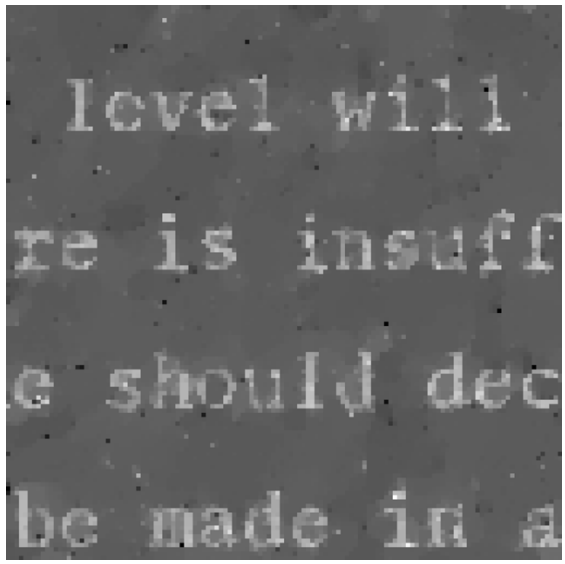


Fig. 13. Standard total variation denoising by (4). Pixel range [−113, 307].

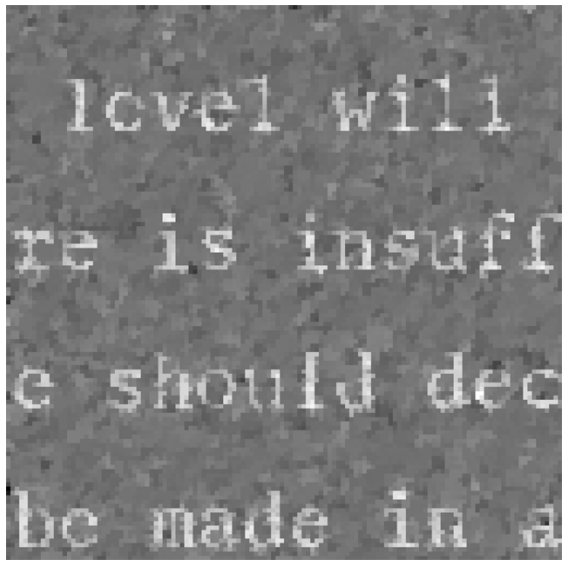


Fig. 14. ℓ^p denoising with bounded total variation. Pixel range [−156, 277].

B. Denoising Application

1) Experiments: Adding i.i.d. Laplacian noise to the original 128×128 gray-level image $\bar{x}$ shown in Fig. 11 results in the image y shown in Fig. 12. The image-to-noise ratio is 1 dB. Here, $\mathcal{H} = \mathbb{R}^N$, where $N = 128^2$.

We first assume knowledge of the noise statistics and no knowledge of the total variation bound. This corresponds to the standard total variation problem (4), whose solution, obtained by the method of [14], is shown in Fig. 13.

A second scenario consists of assuming no probabilistic knowledge of the noise and applying Algorithm 4 with the same constraints sets as in Section IV-A.1. A common objective in denoising problems is the pth power of the ℓ^p norm of the residual, namely

$$J : x \mapsto \|x - y\|_p^p = \sum_{i=0}^{N-1} \left| x^{(i)} - y^{(i)} \right|^p. \tag{21}$$

The impulsive nature of the noise naturally suggests choosing a value of p close to 1. We set $p = 1.1$, which secures that Assumption 2 is satisfied by Proposition 3. The solution to (2), obtained by Algorithm 4 with the total variation constraint set S_1 alone is shown in Fig. 14. The restoration displayed in Fig. 15 is obtained by using the sets $(S_i)_{1 \leq i \leq 3}$ in (2). This experiment illustrates the benefits of added information. The important role played by the total variation constraint set S_1 can be seen in Fig. 16, which shows the image obtained by minimizing J over $S_2 \cap S_3$. The total variation of the images is 4.984×10^5 in Fig. 11, 2.688×10^5 in Fig. 13, 4.996×10^5 in Fig. 14, 4.988×10^5 in Fig. 15, and 6.484×10^5 in Fig. 16.

2) Implementation of Algorithm 4 and Convergence Patterns: Here, we have $x_0 = y$. The main difference with the experiment of Section IV-A.1 is that an explicit solution at Step ⑤ of Algorithm 4 is not available. We can, however, employ a Lagrangian formulation to reduce the problem to the minimization of a convex function of only two nonnegative

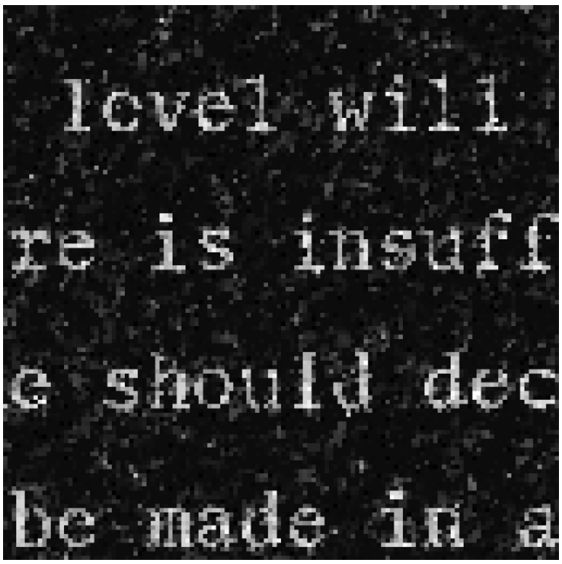


Fig. 15. ℓ^p denoising with bounded total variation and additional constraints. Pixel range $[-1, 255]$.

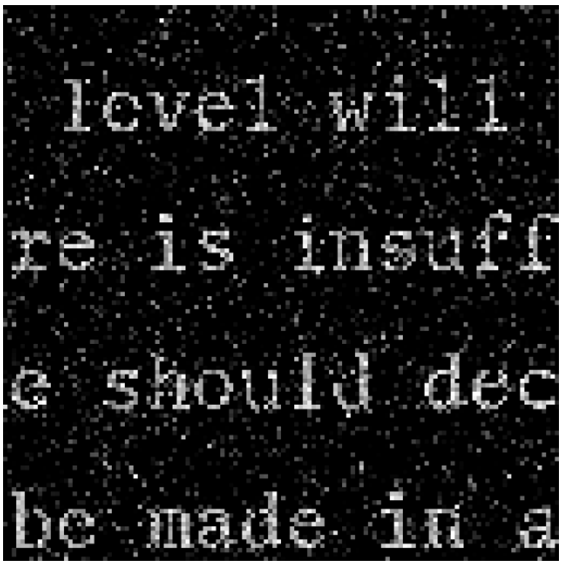


Fig. 16. ℓ^p denoising with no total variation constraint. Pixel range $[-4, 255]$.

real variables. Indeed, standard convex programming results [29, section 28] show that the solution x_{n+1} and the associated Karush–Kuhn–Tucker vector $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2) \in \mathbb{R}^2$ are solutions to the saddle-point problem

$$\max_{\lambda \in \mathbb{R}^2_+} \min_{x \in \mathbb{R}^N} J(x) + \lambda_1 \langle x_n - x \,|\, \nabla J(x_n) \rangle + \lambda_2 \langle x - z_n \,|\, x_n - z_n \rangle. \quad (22)$$

Since the gradient $\nabla J : \mathbb{R}^N \to \mathbb{R}^N$ of the objective is bijective, we can express the solution to the above minimization problem as a function of λ, namely

$$x_\lambda = (\nabla J)^{-1}(\lambda_1 \nabla J(x_n) + \lambda_2 (z_n - x_n)). \quad (23)$$

We then deduce $\bar{\lambda}$ as the solution to the following simple convex minimization problem in $\mathbb{R}^2$

$$\min_{\lambda \in \mathbb{R}^2_+} -J(x_\lambda) + \lambda_1 \langle x_\lambda - x_n \,|\, \nabla J(x_n) \rangle + \lambda_2 \langle x_\lambda - z_n \,|\, z_n - x_n \rangle$$

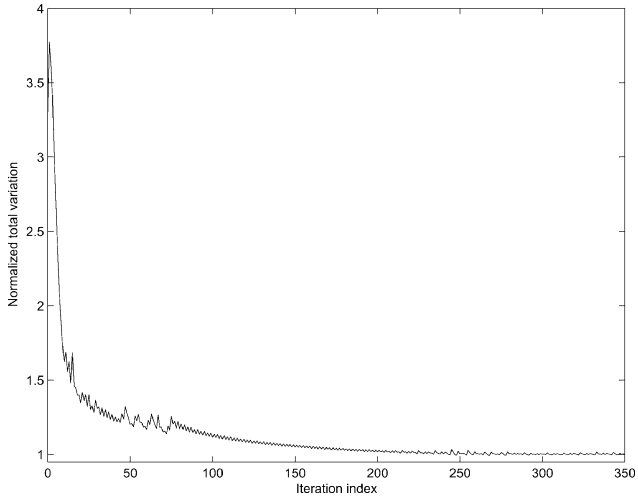


Fig. 17. $\mathrm{tv}(x_n)/\mathrm{tv}(\bar{x})$ versus the iteration index n.

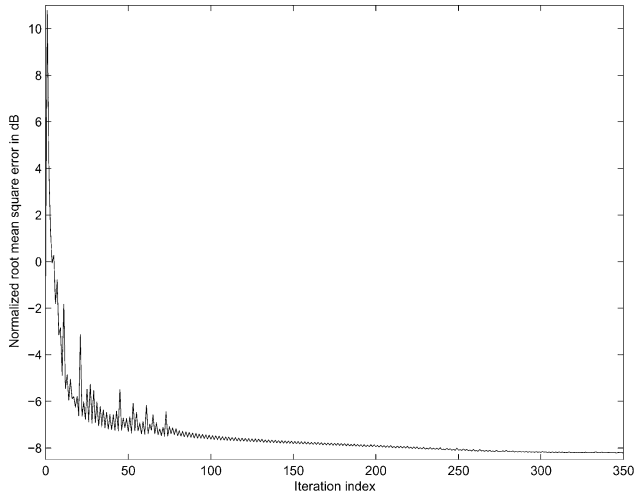


Fig. 18. $\|x_n - \bar{x}\|/\|\bar{x}\|$ versus the iteration index n.

which can be solved by standard numerical methods. In turn, we obtain $x_{n+1} = x_{\bar{\lambda}}$.

The asymptotic behavior of the algorithm in the experiment corresponding to Fig. 15 is illustrated in Figs. 17 and 18, where we show the evolution of the total variation and of the root mean-square error, respectively. Convergence is achieved in roughly 350 iterations.

APPENDIX
APPENDIX–PROOFS

Proof of Proposition 1: First, suppose that $\mathcal{H} = H^1(\Omega)$. Then, tv is defined by (5) and finite on $\mathcal{H}$ (see Section II-C). Now, take x and y in $\mathcal{H}$ and $\alpha \in]0, 1[$. Then, by linearity of $\nabla : \mathcal{H} \to L^2(\Omega) \times L^2(\Omega)$, $|\nabla(\alpha x + (1-\alpha)y)|_2 = |\alpha \nabla x + (1-\alpha)\nabla y|_2 \leq \alpha |\nabla x|_2 + (1-\alpha)|\nabla y|_2$. Hence, $\int_\Omega |\nabla(\alpha x + (1-\alpha)y)|_2 \leq \alpha \int_\Omega |\nabla x|_2 + (1-\alpha) \int_\Omega |\nabla y|_2$, which proves the convexity of tv. Next, since tv is finite and convex, it suffices to show that it is sequentially lower semicontinuous to establish its continuity [17, Cor. I.2.5]. To this end, let $(x_n)_{n \geq 0}$ be a sequence in $\mathcal{H}$ such that $x_n \overset{H^1}{\to} x$ and set $\alpha = \underline{\lim}\, \mathrm{tv}(x_n)$. Then,

we must show that $\mathrm{tv}(x) \leq \alpha$. Let us extract a subsequence $(x_{k_n})_{n\geq 0}$ such that $\mathrm{tv}(x_{k_n}) \to \alpha$. Then

$$\begin{aligned} \|\nabla x_{k_n} - \nabla x\|^2_{L^2\times L^2} &\leq \|x_{k_n} - x\|^2_{L^2} \\ &\quad + \|\nabla x_{k_n} - \nabla x\|^2_{L^2\times L^2} \\ &= \|x_{k_n} - x\|^2_{H^1} \to 0. \end{aligned} \tag{24}$$

Hence, $\nabla x_{k_n} \overset{L^2\times L^2}{\longrightarrow} \nabla x$. Therefore, it follows from [3, theorems 2.5.1 and 2.5.3] that we can extract a further subsequence $(x_{l_{k_n}})_{n\geq 0}$ such that $|\nabla x_{l_{k_n}}|_2 \overset{\text{a.e.}}{\to} |\nabla x|_2$. It now follows from Fatou's lemma [3, Lem. 1.6.8] that

$$\begin{aligned} \mathrm{tv}(x) &= \int_\Omega |\nabla x|_2 = \int_\Omega \lim \left|\nabla x_{l_{k_n}}\right|_2 \\ &\leq \underline{\lim} \int_\Omega \left|\nabla x_{l_{k_n}}\right|_2 = \alpha \end{aligned} \tag{25}$$

which yields the desired inequality. Finally, when $\mathcal{H} = \mathbb{R}^N$, tv is defined by (13) and, by [17, Cor. I.2.3], it is enough to show its convexity. This follows at once from the convexity of the norms and the linearity of the operators involved in (13).

Proof of Proposition 3: Since J is convex and finite, it is continuous [17, Cor. I.2.3] and 2) holds. Furthermore, the coercivity of J implies that all its lower level sets are bounded, so that 3) also holds.

Proof of Theorem 5: The result is an application of [11, theorem 6.4(i)]. First, it follows from [11, (6.10)] and Step ③ c) that H_n at Step ④ of Algorithm 4 can be written as

$$H_n = \left\{ x \in \mathbb{R}^N \;\middle|\; \sum_{i\in I_n} \omega_{i,n} \left\langle x - p_{i,n} \,\middle|\, x_n - p_{i,n} \right\rangle \leq 0 \right\}. \tag{26}$$

Moreover, by construction, $(x_n)_{n\geq 0}$ lies in C [11, Prop. 3.1(ii)]. Hence, in view of [11, (5.4)] and item 3) in Assumption 2, Algorithm 4 is a special case of [11, Algorithm 6.4]. Next, it follows from Assumption 2, [11, Prop. 2.1(i)] and [11, Prop. 2.2(ii)], that assumptions [11, (A1)–(A3)] are satisfied with $E = \mathbb{R}^N$. We conclude by observing that, by [11, Prop. 4.7(ii)] and [29, theorem 24.7], all the conditions of [11, theorem 6.4(i)] are fulfilled.

REFERENCES

[1] A. Abubakar and P. M. van den Berg, "Total variation as a multiplicative constraint for solving inverse problems," *IEEE Trans. Image Processing*, vol. 10, pp. 1384–1392, Sept. 2001.

[2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[3] R. B. Ash, *Real Analysis and Probability*. New York: Academic, 1972.

[4] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, pp. 367–426, Sept. 1996.

[5] M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and H. Azhari, "Reconstruction in diffraction ultrasound tomography using nonuniform FFT," *IEEE Trans. Med. Imag.*, vol. 21, pp. 1395–1401, Nov. 2002.

[6] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, pp. 89–97, Mar. 2004.

[7] A. Chambolle and P. L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.*, vol. 76, pp. 167–188, 1997.

[8] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 22, pp. 503–516, 2000.

[9] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics*, P. Hawkes, Ed. New York: Academic, 1996, vol. 95, pp. 155–270.

[10] ——, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Trans. Image Processing*, vol. 6, pp. 493–506, Apr. 1997.

[11] ——, "Strong convergence of block-iterative outer approximation methods for convex optimization," *SIAM J. Control Optim.*, vol. 38, pp. 538–565, Feb. 2000.

[12] ——, "Convexité et signal," in *Proc. Congrès de Mathématiques Appliquées et Industrielles*, Pompadour, France, May 28–June 1, 2001, pp. 6–16.

[13] ——, "A block-iterative surrogate constraint splitting method for quadratic signal recovery," *IEEE Trans. Signal Processing*, vol. 51, pp. 1771–1782, July 2003.

[14] P. L. Combettes and J. Luo, "An adaptive level set method for nondifferentiable constrained image recovery," *IEEE Trans. Image Processing*, vol. 11, pp. 1295–1304, Nov. 2002.

[15] P. L. Combettes and H. J. Trussell, "The use of noise properties in set theoretic estimation," *IEEE Trans. Signal Processing*, vol. 39, pp. 1630–1641, July 1991.

[16] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2024–2036, Dec. 1989.

[17] I. Ekeland and R. Temam, *Analyse Convexe et Problèmes Variationnels.* Paris: Dunod, 1974; *Convex Analysis and Variational Problems*. Philadelphia, PA: SIAM, 1999.

[18] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*. Boston, MA: Birkhäuser, 1984.

[19] Y. Gousseau and J.-M. Morel, "Are natural images of bounded variation?," *SIAM J. Math. Anal.*, vol. 33, pp. 634–648, 2001.

[20] U. Hermann and D. Noll, "Adaptive image reconstruction using information measures," *SIAM J. Control Optim.*, vol. 38, pp. 1223–1240, Apr. 2000.

[21] B. R. Hunt, "The application of constrained least-squares estimation to image restoration by digital computer," *IEEE Trans. Comput.*, vol. 22, pp. 805–812, Sept. 1973.

[22] S. L. Keeling, "Total variation based convex filters for medical imaging," *Appl. Math. Comput.*, vol. 139, pp. 101–119, 2003.

[23] F. Malgouyres, "Minimizing the total variation under a general convex constraint for image restoration," *IEEE Trans. Image Processing*, vol. 11, pp. 1450–1456, Dec. 2002.

[24] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations—The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*. Providence, RI: AMS, 2001.

[25] M. Nikolova, "Local strong homogeneity of a regularized estimator," *SIAM J. Appl. Math.*, vol. 61, pp. 633–658, 2000.

[26] S. Osher and L. I. Rudin, "Feature-oriented image enhancement using shock filters," *SIAM J. Numer. Anal.*, vol. 27, pp. 919–940, 1990.

[27] I. Pollak, "Segmentation and restoration via nonlinear multiscale filtering," *Signal Processing Mag.*, vol. 19, pp. 26–36, Sept. 2002.

[28] W. Ring, "Structural properties of solutions to total variation regularization problems," *M2AN Math. Model. Numer. Anal.*, vol. 34, pp. 799–810, 2000.

[29] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[30] L. I. Rudin, P. L. Lions, and S. Osher, "Multiplicative denoising and deblurring: Theory and algorithms," in *Geometric Level Set Methods in Imaging, Vision, and Graphics*, S. Osher and N. Paragios, Eds. New York: Springer, 2003, pp. 103–119.

[31] L. I. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Nov. 1994, pp. 31–35.

[32] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.

[33] H. Stark, Ed., *Image Recovery: Theory and Application*. San Diego, CA: Academic, 1987.

[34] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Variational approach for edge-preserving regularization using coupled PDE's," *IEEE Trans. Image Processing*, vol. 7, pp. 387–397, Mar. 1998.

[35] H. J. Trussell, "A priori knowledge in algebraic reconstruction methods," in *Advances in Computer Vision and Image Processing*, T. S. Huang, Ed. Greenwich, CT: JAI, 1984, vol. 1, pp. 265–316.

[36] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 201–212, Apr. 1984.

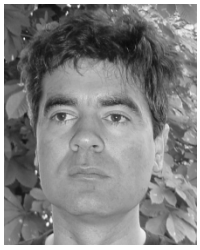
[37] http://images.ee.umist.ac.uk/danny/database.html (UMIST Face Database); see also D. B. Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition" in *Face Recognition: From Theory to Applications*, NATO ASI Series F, Computer and Systems Sciences, vol. 163, H. Wechsler, J. P. Phillips, V. Bruce, F. Fogelman-Soulié, and T. S. Huang (Eds.), pp. 446–456, 1998.

[38] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Processing*, vol. 7, pp. 813–824, June 1998.

[39] J. Weickert, *Anisotropic Diffusion in Image Processing*. Stuttgart, Germany: Teubner-Verlag, 1998.

[40] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," *IEEE Trans. Med. Imag.*, vol. 1, pp. 81–94, Oct. 1982.

[41] W. P. Ziemer, *Weakly Differentiable Functions*. New York: Springer, 1989.



Patrick Louis Combettes (S'84–M'90–SM'96) is a Professor with the Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie - Paris 6, Paris, France.

Dr. Combettes is the recipient of the IEEE Signal Processing Society Paper Award.

Jean-Christophe Pesquet (S'89–M'91–SM'99) received the engineering degree from Supélec, Gif-sur-Yvette, France, in 1987, the Ph.D. degree from the Université Paris-Sud (XI), Paris, France, in 1990, and the Habilitation à Diriger des Recherches from the Université Paris-Sud in 1999.

From 1991 to 1999, he was a Maître de Conférences with the Université Paris-Sud and a Research Scientist at the Laboratoire des Signaux et Systèmes, Centre National de la Recherche Scientifique (CNRS), Gif-sur-Yvette. He is currently a Professor with Université de Marne-la-Vallée, France, and a Research Scientist at the Laboratoire d'Informatique of the university (UMR-CNRS 8049), Paris.