1 2 3 4

An Adaptive Level Set Method for Nondifferentiable Constrained Image Recovery

Patrick L. Combettes, *Senior Member, IEEE* and Jian Luo

Abstract—**The formulation of a wide variety of image recovery problems leads to the minimization of a convex objective over a convex set representing the constraints derived from *a priori* knowledge and consistency with the observed signals. In recent years, nondifferentiable objectives have become popular due in part to their ability to capture certain features such as sharp edges. They also arise naturally in minimax inconsistent set theoretic recovery problems. At the same time, the issue of developing reliable numerical algorithms to solve such convex programs in the context of image recovery applications has received little attention. In this paper, we address this issue and propose an adaptive level set method for nondifferentiable constrained image recovery. The asymptotic properties of the method are analyzed and its implementation is discussed. Numerical experiments illustrate applications to total variation and minimax set theoretic image restoration and denoising problems.**

Keywords— **Image recovery, level set method, nondifferentiable optimization, reconstruction, restoration, total variation.**

I. Introduction

A broad range of digital image restoration, reconstruction, and denoising problems can be formulated as constrained convex optimization problems of the form

$$\text{Find } \; x^* \in S \; \text{ such that } \; J(x^*) = \inf J(S), \tag{1}$$

where S is a closed convex set in the standard N-dimensional Euclidean space $\mathbb{R}^N$ describing image constraints derived from *a priori* knowledge and consistency with the observed signals, and $J\colon \mathbb{R}^N \to \mathbb{R}$ is a convex function. Typically, the feasibility set S represents information known *a priori* about the image to be recovered and the physical system that generated the measured data [11], [14], [37], [39], [43], while J allows for the selection of an image in the feasibility set [5], [12], [14], [27], [29], [31], [35].

The relative ease of implementation of smooth minimization methods has traditionally favored the use of differentiable objectives in (1), e.g., [5], [9], [12], [27], [38]. In recent years, however, it has emerged from various theoretical and experimental studies that nondifferentiable objectives were more appropriate in certain signal and image recovery problems, due in part to their ability to restitute sharp features [3], [4], [10], [23], [29], [35], [41]. As will be seen in Section V, nondifferentiable objectives also arise naturally in minimax formulations for inconsistent set theoretic image recovery problems. At the same time, there has been limited activity towards the design of reliable numerical algorithms for solving the nondifferentiable convex program (1) in the context of image recovery applications.

In nonsmooth optimization problems, gradients may not be defined and the usual recourse is to use subgradients. Unfortunately the latter contain much less information than the former and, for that reason, nonsmooth minimization problems must be tackled with specific algorithms. While it may be tempting to just employ a smooth optimization algorithm to solve (1) with a nondifferentiable objective, such a practice should be strongly discouraged as it may lead to dramatic failures [24], [28], [36]. An alternative is to approximate J in (1) by a smooth function and to employ a smooth minimization scheme to solve the approximate problem (this approach was adopted in the total variation problems of [10], [41]). Although conceptually simple, this smooth approximation approach has three serious shortcomings:

- There is no systematic procedure to construct smooth approximations to nondifferentiable functions.
- By smoothing the original objective, one forfeits the theoretical justification that precisely led to the selection of a nondifferentiable cost function since it is in general unclear how well a solution to the perturbed problem approximates, in a physically meaningful sense, those of the exact problem.
- A good smooth approximation to a nondifferentiable function is "stiff", i.e., its gradient varies continuously but rapidly. As demonstrated in [24, Section VIII.3.3, Vol. I], stiff functions are hard to minimize via smooth optimization techniques and should actually be handled as nondifferentiable functions. For instance, the range of the step-size of the projected gradient method for solving (1) with a κ-Lipschitz objective is bounded by $2/\kappa$ [7, Prop. 3.3.4]. As κ is large for stiff functions, the method is unviable numerically.[1] More technical pitfalls of smooth approximation techniques are discussed in [30] in the context of phase recovery problems.

Other alternatives have been explored in specific signal recovery problems. For instance, in the quadratically constrained total variation image denoising problem of [35], the scheme which is used is akin to a projected gradient method in which iterates are perturbed to avoid points of nondifferentiability. This heuristic approach is straightforward to implement but lacks a sound mathematical basis. In the quadratically constrained image restoration problem of [29], a variant of the total variation objective led to an ℓ_1 problem that was solved by an affine scaling Newton method whose computational load is a handicap for large images. It should also be noted that several standard nonsmooth minimization methods are practical only in small-scale

Manuscript received XX X, 2001; revised XX X, 2002. This work was supported in part by the National Science Foundation under grant MIP-9705504. The associate editor coordinating the review of this paper and approving it for publication was Dr. T. Schulz.

P. L. Combettes is with the Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 4 place Jussieu, 75005 Paris, France (e-mail plc@math.jussieu.fr).

J. Luo is with the Department of Electrical Engineering, City College and Graduate Center, City University of New York, New York, NY 10031, USA (e-mail ljian@ee-mail.engr.ccny.cuny.edu).

Publisher Item Identifier XXXX.

[1] As a simple illustration, consider the smooth approximation $J_\varepsilon\colon x \mapsto \sqrt{\|x\|^2+\varepsilon}$ to the nondifferentiable function $J\colon x \mapsto \|x\|$ for a small parameter $\varepsilon > 0$. The Lipschitz constant of J_ε is $1/\sqrt{\varepsilon}$ and step-sizes must therefore be less than $2\sqrt{\varepsilon}$.

problems [24], [36] and are therefore ruled out in image recovery applications.

The goal of this paper is to propose an implementable, practical, and reliable algorithm for solving the constrained image recovery problem (1) with nondifferentiable objectives. The principle of the proposed adaptive level set algorithm is common to several state-of-the-art schemes in nonsmooth optimization, e.g., [25], [26], that have evolved from Polyak's projected subgradient method [32]. Unlike the methods currently in use in image recovery, only mild assumptions on the objective J (convexity) and the constraint set S (convexity and compactness) are required, and (1) is solved without being altered. As a result, the algorithm is applicable to a wide range of recovery problems. In addition, several aspects of the practical implementation of the algorithm are discussed, with special emphasis on stopping rules.

In Section II, the necessary mathematical background is briefly reviewed. We then describe Polyak's method itself as well as two variants that will be essential ingredients in the design of our algorithm. In Section III, we present the algorithm, establish its convergence, and discuss its implementation. We report on numerical applications to total variation image restoration and denoising in Section IV and to minimax set theoretic image restoration in Section V. Appendix A contains the proofs of technical results.

II. Mathematical foundation

A. Basic facts

We recall here some basic facts; details can be found in [24], [33], [34].

Let $J\colon \mathbb{R}^N \to \mathbb{R}$ be a convex function. Then J is continuous on $\mathbb{R}^N$ and, for every $\alpha \in \mathbb{R}$, its *lower level set* at height α, $\operatorname{lev}_{\leq\alpha} J = \big\{x \in \mathbb{R}^N \mid J(x) \leq \alpha\big\}$, is closed and convex. A vector $t \in \mathbb{R}^N$ is a *subgradient* of J at $x \in \mathbb{R}^N$ if $(\forall y \in \mathbb{R}^N)$ $\langle y - x \mid t\rangle + J(x) \leq J(y)$. The set of all subgradients of J at x is the *subdifferential* of J at x; it is nonempty and denoted by $\partial J(x)$. If J is differentiable at x, then its gradient $\nabla J(x)$ is its unique subgradient: $\partial J(x) = \{\nabla J(x)\}$. In addition, $x \in \mathbb{R}^N$ is a global minimizer of J if and only if $0 \in \partial J(x)$.

Now let C be a nonempty closed convex set in $\mathbb{R}^N$. Then, for every $x \in \mathbb{R}^N$, there exists a unique point $P_C(x) \in C$ such that $\|x - P_C(x)\| = d(x, C) = \inf_{y\in C} \|x - y\|$. The point $P_C(x)$ is the *projection* of x onto C. The projector $P_C\colon \mathbb{R}^N \to C$ satisfies

$$(\forall x \in \mathbb{R}^N)(\forall y \in C)\ \ \|P_C(x)-y\|^2 \leq \|x-y\|^2 - \|P_C(x)-x\|^2 \tag{2}$$

and the distance function $d(\cdot, C)$ is convex and differentiable at every point $x \in \mathbb{R}^N \smallsetminus C$ with

$$\nabla d(x, C) = \frac{x - P_C(x)}{d(x, C)}. \tag{3}$$

B. Subgradient projection

A tutorial account of subgradient projections can be found in [15].

Let $J\colon \mathbb{R}^N \to \mathbb{R}$ be a convex function, α a real number, and g a *selection* of ∂J, i.e., $(\forall x \in \mathbb{R}^N)$ $g(x) \in \partial J(x)$. Then

$$\operatorname{lev}_{\leq\alpha} J \subset H^g_\alpha(x) = \big\{y \in \mathbb{R}^N \mid \langle x - y \mid g(x)\rangle \geq J(x) - \alpha\big\}.$$

The *subgradient projection* $G^g_\alpha(x)$ of x onto $\operatorname{lev}_{\leq\alpha} J$ is

$$G^g_\alpha(x) = \begin{cases} x & \text{if } J(x) \leq \alpha \ \text{ or } \ g(x) = 0 \\ x - \dfrac{J(x) - \alpha}{\|g(x)\|^2}\, g(x) & \text{otherwise.} \end{cases} \tag{4}$$

As seen above, the situation $g(x) = 0$ may occur only when x is a global minimizer of J. If $\operatorname{lev}_{\leq\alpha} J \neq \varnothing$, then $G^g_\alpha(x)$ is the projection of x onto the closed halfspace $H^g_\alpha(x)$. Since the computation of $G^g_\alpha(x)$ requires only a subgradient $g(x)$ (the gradient $\nabla J(x)$, in the differentiable case) of J at x, subgradient projections are significantly easier to implement than exact projections and have been used for solving a wide range of feasibility problems [6], [14], [15]. For subsequent use, we record the fact that subgradient projections satisfy a property akin to (2), namely

$$(\forall x \in \mathbb{R}^N)(\forall y \in \operatorname{lev}_{\leq\alpha} J)\ \ \|G^g_\alpha(x) - y\|^2 \leq \|x - y\|^2 - \|G^g_\alpha(x) - x\|^2. \tag{5}$$

C. Standing assumptions

Throughout the paper our assumptions regarding problem (1) are as follows. $J\colon \mathbb{R}^N \to \mathbb{R}$ is a convex function and S is a nonempty compact convex subset of $\mathbb{R}^N$. Consequently, $\alpha^* = \inf J(S) > -\infty$ and the solution set $S^* = S \cap \operatorname{lev}_{\leq\alpha^*} J$ is compact, convex, and nonempty [33]. In addition, g designates an arbitrary selection of ∂J.

D. Polyak's subgradient projection method

The subgradient projection method is governed by the iterative process

$$x_0 \in S \ \text{ and } \ (\forall n \in \mathbb{N}) \ \begin{cases} x_{n+1} = P_S\left(x_n - \sigma_n \dfrac{g(x_n)}{\|g(x_n)\|}\right), \\ \sigma_n > 0. \end{cases} \tag{6}$$

For this algorithm, a typical convergence condition on the step-sizes $(\sigma_n)_{n\geq 0}$ is $\sum_{n\geq 0} \sigma_n = +\infty$ and $\sum_{n\geq 0} \sigma_n^2 < +\infty$, e.g., [33]. This condition implies that the step-sizes must converge rapidly to zero, which translates into slow convergence. To circumvent this problem, Polyak proposed a different type of step-sizes under the assumption that the optimal value $\alpha^* = \inf J(S)$ is known [32], [36]. He showed that a sequence $(x_n)_{n\geq 0}$ converging to some $x \in S^*$ can be generated by (6) where $(\forall n \in \mathbb{N})$ $\sigma_n = (J(x_n) - \alpha^*)/\|g(x_n)\|$. With these step-sizes, (6) becomes

$$x_0 \in S \ \text{ and } \ (\forall n \in \mathbb{N}) \ x_{n+1} = P_S \circ G^g_{\alpha^*}(x_n) \tag{7}$$

Polyak's algorithm consists in alternating a subgradient projection onto $\operatorname{lev}_{\leq\alpha^*} J$ and an exact projection onto S and is therefore a special case of the general subgradient projection schemes of [6], [15]. Unfortunately, it is implementable only in those rare instances when α^* is known.

When α^* is unknown, a general strategy is to replace (7) by the adaptive level set method

$$x_0 \in S \ \text{ and } \ (\forall n \in \mathbb{N}) \ x_{n+1} = P_S \circ G^g_{\alpha_n}(x_n), \tag{8}$$

where $(\alpha_n)_{n\geq 0}$ is a sequence of level estimates [2], [25], [26] (*cf.* Fig. 1). If an upper bound α on α^* is available, an approximate solution to (1) can be constructed as follows.

Theorem 1 [25] *Suppose $\alpha \geq \alpha^*$ and $x_0 \in S$. Let $(x_n)_{n\geq 0}$ be a sequence generated by* (8) *with $\alpha_n \downarrow \alpha$. Then $(x_n)_{n\geq 0}$ converges to a point in $S \cap \operatorname{lev}_{\leq\alpha} J$.*

Now suppose that we know a lower bound α on α^*. Then, as $S \cap \operatorname{lev}_{\leq\alpha} J = \varnothing$, a sequence generated by (8) with $\alpha_n \equiv \alpha$ cannot converge. It nonetheless possesses a property that will prove very useful.

Theorem 2 [2] *Suppose $\alpha < \alpha^*$ and $x_0 \in S$. Let $(x_n)_{n\geq 0}$ be a sequence generated by* (8) *with $(\forall n \in \mathbb{N})\ \alpha_n = \alpha$. Then $\big(\forall \varepsilon \in \,]0,+\infty[\big)(\exists\, m \in \mathbb{N})\ J(x_m) \leq \alpha^* + (\alpha^* - \alpha) + \varepsilon$.*

Theorems 1 and 2 imply that if a reasonably tight estimate of α^* is available then (1) can be solved approximately by (8). While these two theorems can usually not be used directly in practice for lack of a good approximation to α^*, they describe general principles that constitute the foundation of the algorithms presented in [20], [25], [26] and of the algorithm proposed in this paper.

III. Algorithm

A. Description

Adaptive level set methods are based on the following principle. Let α be a guess of the optimal level value α^*. Then (*cf.* Fig. 2):

- If $x \in S \cap \operatorname{lev}_{\leq\alpha} J$ can be found, we infer that $S \cap \operatorname{lev}_{\leq\alpha} J \neq \varnothing$ and therefore that $\alpha \geq \alpha^*$.
- If $S \cap \operatorname{lev}_{\leq\alpha} J = \varnothing$ is detected, we infer that $\alpha < \alpha^*$.

Theorems 1 and 2 can be exploited to transform (8) into an adaptive level set method. First note that, since $(x_n)_{n\geq 0}$ lies in S, $\inf_{n\geq 0} J(x_n) \geq \alpha^*$. Therefore, if we define

$$\overline{\alpha}_0 = J(x_0) \ \text{ and } \ (\forall n \in \mathbb{N})\ \overline{\alpha}_{n+1} = \min\{J(x_{n+1}), \overline{\alpha}_n\}, \tag{9}$$

then

$$(\forall n \in \mathbb{N})\ \overline{\alpha}_n \geq \overline{\alpha}_{n+1} = \min_{0\leq m\leq n+1} J(x_m) \geq \alpha^*. \tag{10}$$

Now fix $0 < \lambda < 1$, $\varepsilon > 0$, and $\eta_0 > \lambda\varepsilon$. Define two sequences $(\alpha_n)_{n\geq 0}$ and $(\eta_n)_{n\geq 0}$ in $\mathbb{R}$ by

$$(\forall n \in \mathbb{N})\ \ \alpha_n = \overline{\alpha}_n - \eta_n \ \text{ and}$$
$$\eta_{n+1} = \begin{cases} \eta_n & \text{if } S \cap \operatorname{lev}_{\leq\alpha_n} J \neq \varnothing \\ \eta_n & \text{if } S \cap \operatorname{lev}_{\leq\alpha_n} J = \varnothing \text{ is not detected} \\ \lambda\eta_n & \text{if } S \cap \operatorname{lev}_{\leq\alpha_n} J = \varnothing \text{ is detected.} \end{cases} \tag{11}$$

Then

$$(\forall n \in \mathbb{N})\ \begin{cases} \eta_{n+1} \leq \eta_n \\ (\exists\, m \in \{0,\ldots,n\})\ \ \eta_n = \lambda^m \eta_0. \end{cases} \tag{12}$$

As we shall see, with such a construction, $\overline{\alpha}_n$ approaches α^* from above whereas η_n approaches 0 from above. Whence, α_n approaches α^*. In view of (10)–(11) and the fact that the update $\eta_{n+1} = \lambda\eta_n$ takes place only if infeasibility $S \cap \operatorname{lev}_{\leq\alpha_n} J = \varnothing$ ($\alpha_n < \alpha^*$) is detected, the occurrence of the inequality $\eta_n \leq \lambda\varepsilon$ can be used as a termination criterion, where $\varepsilon > 0$ is a preset tolerance on α^* (*cf.* proof of Theorem 4). As seen in Section II-A, if $g(x_n) = 0$, then x_n is a global minimizer of J and *a fortiori* a solution to (1), which justifies using $g(x_n) = 0$ as a second stopping rule. On the other hand, if $g(x_n) \neq 0$, since $J(x_n) > \alpha_n$ by (9) and (11), (4) yields $G^g_{\alpha_n}(x_n) = x_n + (\alpha_n - J(x_n))g(x_n)/\|g(x_n)\|^2$. These considerations lead to the following conceptual algorithm.

Algorithm 3 Fix $v \in \mathbb{R}^N$, $\varepsilon > 0$, and $0 < \lambda < 1$.

Step 0. Set $x_0 = P_S(v)$, $\eta_0 > \lambda\varepsilon$, $\overline{\alpha}_0 = J(x_0)$, and $n = 0$.
Step 1. If $\eta_n \leq \lambda\varepsilon$, terminate.
Step 2. Obtain $g_n \in \partial J(x_n)$. If $g_n = 0$, terminate.
Step 3. Set $\alpha_n = \overline{\alpha}_n - \eta_n$.
Step 4. Set $x_{n+1} = P_S\big(x_n + (\alpha_n - J(x_n))g_n/\|g_n\|^2\big)$.
Step 5. If $S \cap \operatorname{lev}_{\leq\alpha_n} J = \varnothing$ is detected, go to Step 6; Otherwise, go to Step 7.
Step 6. Set $\eta_{n+1} = \lambda\eta_n$, $\overline{\alpha}_{n+1} = \overline{\alpha}_n$, $x_{n+1} = x_n$, $n = n+1$, and go to Step 1.
Step 7. Set $\eta_{n+1} = \eta_n$, $\overline{\alpha}_{n+1} = \min\{J(x_{n+1}), \overline{\alpha}_n\}$, $n = n+1$, and go to Step 2.

The basic mechanism to refine adaptively the approximate levels $(\alpha_n)_n$ is the following. At iteration n, $\overline{\alpha}_n \geq \alpha^*$ is available and we construct the new estimate α_n by decreasing $\overline{\alpha}_n$ by a factor $\eta_n > 0$ (Step 3). If it can be detected that $S \cap \operatorname{lev}_{\leq\alpha_n} J$ is empty, then we deduce that $\alpha_n < \alpha^*$ and therefore that η_n is too large. We then scale η_n down by a factor λ (Step 6) and re-execute this loop with this smaller value of η_n. Otherwise, we update $\overline{\alpha}_n$ to a lower value (Step 7) and rerun the loop with the same value of η_n.

It is important to note that if infeasibility occurs at iteration n (i.e., $S \cap \operatorname{lev}_{\leq\alpha_n} J = \varnothing$ or, equivalently, $\alpha_n < \alpha^*$) but is not detected, then α_n is updated to a value $\alpha_{n+1} \leq \alpha_n$ since

$$\begin{aligned} \alpha_{n+1} = \overline{\alpha}_{n+1} - \eta_{n+1} &= \min\{J(x_{n+1}), \overline{\alpha}_n\} - \eta_n \\ &\leq \overline{\alpha}_n - \eta_n = \alpha_n. \end{aligned} \tag{13}$$

Thus, if we define the infeasibility gap at iteration n by $\delta_n = \alpha^* - \alpha_n$, any undetected infeasibility leads to another infeasibility with a gap δ_{n+1} at least as wide. Our basic premise is that every infeasibility can be eventually detected in the sense that

$$\text{If } \ S \cap \operatorname{lev}_{\leq\alpha_n} J = \varnothing, \ \text{ then} \quad (\exists\, k \in \mathbb{N})\ \ S \cap \operatorname{lev}_{\leq\alpha_{n+k}} J = \varnothing \ \text{ is detected.} \tag{14}$$

It will be justified by concrete detection rules in Section III-C.1.

B. Main result

Our main result states that Algorithm 3 produces a signal in S that satisfies any preset tolerance on the constrained objective, i.e., an approximate solution to (1) that is feasible and can be made arbitrarily close to optimal.

Theorem 4 *Fix $\varepsilon > 0$. Then, under assumption* (14)*, Algorithm 3 generates a point x_n in S such that $J(x_n) \leq \alpha^* + \varepsilon$.*

C. Implementation

The implementation of Algorithm 3 is straightforward except for the infeasibility detection condition (14) and the computation of the projection onto S. We now address these issues.

C.1 Infeasibility detection

Given $\alpha_n < \alpha^*$, the problem is to devise a numerical scheme to detect $S \cap \operatorname{lev}_{\leq \alpha_{n+k}} J = \varnothing$ for some $k \in \mathbb{N}$. To this end define, for every $m \in \mathbb{N}$, $\mathbb{N}_m = \{n \in \mathbb{N} \mid \eta_n = \lambda^m \eta_0\}$. In other words, $\mathbb{N}_m$ is the interval of iteration indices over which the parameter η_n is not updated and kept at value $\lambda^m \eta_0$ because infeasibility is not detected (whether or not it actually occurs). We shall denote by l_m the smallest integer in $\mathbb{N}_m$, i.e., the index of the iteration of the mth infeasibility detection. We also need to define (*cf.* Fig. 1 for a geometric interpretation)

$$(\forall n \in \mathbb{N}_m) \;\; \rho_n = \|G^g_{\alpha_n}(x_n) - x_n\|^2 + \|P_S \circ G^g_{\alpha_n}(x_n) - G^g_{\alpha_n}(x_n)\|^2. \quad (15)$$

Proposition 5 *For every $m \in \mathbb{N}$, the following holds.*
(i) $(\alpha_n)_{n \in \mathbb{N}_m}$ *is nonincreasing and* $(\exists\, n \in \mathbb{N}_m)\; \alpha_n < \alpha^*$.
(ii) *If $\mathbb{N}_m$ is infinite, then* $\sum_{n \in \mathbb{N}_m} \rho_n = +\infty$.
(iii) *For every $n \in \mathbb{N}_m$ and $\gamma_m \geq d(x_{l_m}, S^*)$, if*

$$\sum_{k=l_m}^{n} \rho_k > \|x_{l_m} - x_{n+1}\| \big(2\gamma_m - \|x_{l_m} - x_{n+1}\|\big), \quad (16)$$

then $S \cap \operatorname{lev}_{\leq \alpha_n} J = \varnothing$.

Item (i) asserts that

$$(\forall m \in \mathbb{N})(\exists\, n \in \mathbb{N}_m) \;\; S \cap \operatorname{lev}_{\leq \alpha_n} J = \varnothing \quad (17)$$

and, consequently, that infeasibility does occur if η_n is kept constant. Moreover, the accumulation of the squared steps $\sum_{k=l_m}^{n} \rho_k$ grows indefinitely large as n increases (item (ii)) and signals infeasibility precisely when it exceeds the value $\|x_{l_m} - x_{n+1}\|(2\gamma_m - \|x_{l_m} - x_{n+1}\|)$ (item (iii)). This result provides us with a practical detection rule to implement Step 5 of Algorithm 3. Naturally, our goal is to detect infeasibility as soon as possible after it occurs, which means that the parameter γ_m in (16) should be a tight approximation to $d(x_{l_m}, S^*)$. Since $d(x_{l_m}, S^*) = \|P_{S^*}(x_{l_m}) - x_{l_m}\| \leq \sup_{(x,y) \in S^2} \|x - y\| = \operatorname{diam} S$, a conservative choice for γ_m is to use an upper bound on the diameter of S. In most problems, however, one will be able to use tighter estimates of $d(x_{l_m}, S^*)$ based on prior experience and theoretical or heuristic considerations. For instance, if J is strongly convex, a tight bound γ_m can be derived from the results of [25].

The practical detection rule (16) leads to the following implementable version of Algorithm 3.

Algorithm 6 Fix $v \in \mathbb{R}^N$, $\varepsilon > 0$, and $0 < \lambda < 1$.
Step 0. Set $x_0 = P_S(v)$, $\eta_0 > \lambda\varepsilon$, $\overline{\alpha}_0 = J(x_0)$, and $n = 0$.
Step 1. Set $\rho = 0$, $y = x_n$, and $\gamma \geq d(y, S^*)$. If $\eta_n \leq \lambda\varepsilon$, terminate.
Step 2. Obtain $g_n \in \partial J(x_n)$. If $g_n = 0$, terminate.
Step 3. Set $\alpha_n = \overline{\alpha}_n - \eta_n$.
Step 4. Set $w = x_n + (\alpha_n - J(x_n)) g_n / \|g_n\|^2$, $x_{n+1} = P_S(w)$, and $\rho = \rho + \|w - x_n\|^2 + \|x_{n+1} - w\|^2$.
Step 5. Set $\beta = \|y - x_{n+1}\|$. If $\rho > \beta(2\gamma - \beta)$ go to Step 6; Otherwise, go to Step 7.
Step 6. Set $\eta_{n+1} = \lambda\eta_n$, $\overline{\alpha}_{n+1} = \overline{\alpha}_n$, $x_{n+1} = x_n$, $n = n+1$, and go to Step 1.
Step 7. Set $\eta_{n+1} = \eta_n$, $\overline{\alpha}_{n+1} = \min\{J(x_{n+1}), \overline{\alpha}_n\}$, $n = n + 1$, and go to Step 2.

C.2 Projection onto S

As with any variant of the projected subgradient algorithm, the performance of our algorithm is sensitive to the cost of computing the projection onto the feasibility set S at Step 4. If S is derived from a single constraint, the projection onto it is often known in closed form (see [14], [37], [43] for standard examples). On the other hand, if S is specified as an intersection of closed convex sets $(C_i)_{1 \leq i \leq r}$, the projection problem must be decomposed into elementary problems relative to each C_i. Several iterative methods of this type were reviewed in [12], which require only the ability to project onto each set C_i individually and have essentially the same numerical complexity as the cyclic projection (POCS) method of [8] (see also [43]). When the projectors onto the individual sets $(C_i)_{1 \leq i \leq r}$ cannot be implemented in a straightforward fashion, these methods may be demanding numerically and one should turn to the method recently proposed in [16, Section 6.5], which requires only subgradient projections and can therefore construct the projection onto S quite efficiently.

D. Comparisons with existing level set methods

Although the general structure of Algorithm 3 is akin to that of those presented in [20], [25], [26], it differs from these algorithms in several respects. In the algorithm proposed in [20], the level α_n is of the form (11) and $\alpha^* = -\infty$ is allowed. However, since the update of η_n is not based on infeasibility detection as in Algorithm 3, it is not clear how to devise a tractable termination rule. On the other hand, the algorithm proposed in [25] features a different scheme to implement infeasibility detection at Step 5. Finally, in [26], instead of using the subgradient projection $G^g_{\alpha_n}(x_n)$ at Step 4, the projections onto successive approximations to $\operatorname{lev}_{\leq \alpha_n} J$ derived from several accumulated subgradient of J or their aggregates are used. We emphasize that, since

$$\begin{cases} \sum_{k=l_m}^{n} \rho_k \geq \sum_{k=l_m}^{n} \|G^g_{\alpha_k}(x_k) - x_k\|^2 \\ \gamma_m^2 \geq \|x_{l_m} - x_{n+1}\| \big(2\gamma_m - \|x_{l_m} - x_{n+1}\|\big), \end{cases} \quad (18)$$

our detection rule (16) is tighter than that of [25], namely, $\sum_{k=l_m}^{n} \|G^g_{\alpha_k}(x_k) - x_k\|^2 > \gamma_m^2$. It is also tighter than that of [26], namely, $\sum_{k=l_m}^{n} \rho_k > \zeta$, where $\zeta \geq (\operatorname{diam} S)^2 \;\big(\geq \gamma_m^2\big)$.

IV. Application to Total Variation Image Recovery

A. Total variation

Under suitable assumptions (*cf.* [19] for theoretical details), the total variation of a real-valued function x defined on a smooth open subset $\Omega \subset \mathbb{R}^2$ is

$$J_{\rm tv}(\mathsf{x}) = \int_\Omega |\nabla \mathsf{x}(\omega)|_2 d\omega, \tag{19}$$

where $|\cdot|_2$ denotes the Euclidean norm in $\mathbb{R}^2$. This function has been proposed in [35] as an optimality criterion for image denoising and then used as an optimality criterion for image restoration, e.g., [1], [10], [29], [41]. The motivation for minimizing $J_{\rm tv}$ in such problems lies in that it does not penalize discontinuities and tends to preserve the location of the edges of the original image. It is therefore appropriate for piecewise smooth images and, in particular, for images that have block features [10], [35], [41].

Now consider a compactly supported two-dimensional image x which has been discretized on an $M \times M$ grid. The total variation of the discretized image matrix $x \in \mathbb{R}^{M\times M}$ can be obtained through the approximations

$$\begin{cases} \mathsf{x}(\omega) \to x^{i,j} \\ |\nabla \mathsf{x}(\omega)|_2 \to \sqrt{|x^{i+1,j} - x^{i,j}|^2 + |x^{i,j+1} - x^{i,j}|^2} \\ \int \to \sum, \end{cases} \tag{20}$$

where $x^{i,j}$ denotes the (i,j)th pixel of x. Taking into account boundary effects, the total variation of x is defined as

$$\begin{aligned} J_{\rm tv}(x) \;=\; & \sum_{i=0}^{M-2}\sum_{j=0}^{M-2} \sqrt{|x^{i+1,j} - x^{i,j}|^2 + |x^{i,j+1} - x^{i,j}|^2} \\ & + \sum_{i=0}^{M-2} |x^{i+1,M-1} - x^{i,M-1}| \\ & + \sum_{j=0}^{M-2} |x^{M-1,j+1} - x^{M-1,j}|. \end{aligned} \tag{21}$$

To study the properties of $J_{\rm tv}$ it is more convenient to employ the usual column stacking isometry $x^{i,j} \leftrightarrow x^{i+Mj}$ [5] and deal with x as a vector in $\mathbb{R}^N$, where $N = M^2$. In turn, upon introducing suitable difference matrices $(L_{i,j})_{0\le i,j\le M-2}$ in $\mathbb{R}^{2\times N}$, $(L_{i,M-1})_{0\le i\le M-2}$ in $\mathbb{R}^{1\times N}$, and $(L_{M-1,j})_{0\le j\le M-2}$ in $\mathbb{R}^{1\times N}$, the total variation of $x \in \mathbb{R}^N$ is given by

$$J_{\rm tv}(x) = \sum_{i=0}^{M-2}\sum_{j=0}^{M-2} |L_{i,j}x|_2 + \sum_{i=0}^{M-2} |L_{i,M-1}x| + \sum_{j=0}^{M-2} |L_{M-1,j}x|. \tag{22}$$

Since the norms are convex, the composition of the norms with linear operators is also convex and so is their sum. This shows that $J_{\rm tv}$ is a convex function. Let us now turn to its differentiability properties. Since the Euclidean norm is differentiable except at the origin, $J_{\rm tv}$ is not differentiable at x if any of the following conditions holds:

(a) For some $0 \le i,j \le M-2$, $L_{i,j}x = 0$, i.e., pixel-wise, $x^{i+1,j} = x^{i,j} = x^{i,j+1}$.

(b) For some $0 \le i \le M-2$, $L_{i,M-1}x = 0$, i.e., pixel-wise, $x^{i+1,M-1} = x^{i,M-1}$.

(c) For some $0 \le j \le M-2$, $L_{M-1,j}x = 0$, i.e., pixel-wise, $x^{M-1,j+1} = x^{M-1,j}$.

Step 2 of Algorithm 6 requires the computation of a subgradient g_n of $J_{\rm tv}$ at x_n. By the subdifferential sum theorem [24, Thm. VI.4.1.1], it suffices to add subgradients for each of the individual terms making up the sum in (22). Regarding the contribution of the term $|L_{i,j}x_n|_2$ to g_n, there are two alternatives:

- If (a) holds, $x \mapsto |L_{i,j}x|_2$ is nondifferentiable and its subdifferential at x_n is given by $\partial|L_{i,j}x_n|_2 = L_{i,j}^T\big(B_2(0;1)\big)$, where $B_2(0;1)$ is the closed unit disk in $\mathbb{R}^2$ [34, Section 23]. Since $0 \in \partial|L_{i,j}x_n|_2$, we may elect to ignore the contribution of this term.
- If (a) does not hold, $x \mapsto |L_{i,j}x|_2$ is differentiable at x_n and $\nabla|L_{i,j}x_n|_2 = L_{i,j}^T L_{i,j}x_n/|L_{i,j}x_n|_2$ is its unique subgradient.

A term of the form $|L_{i,M-1}x_n|$ (resp. $|L_{M-1,j}x_n|$) can be treated similarly: if condition (b) (resp. (c)) holds, 0 is an acceptable subgradient and the contribution of $|L_{i,M-1}x_n|$ (resp. $|L_{M-1,j}x_n|$) may therefore be ignored; otherwise, the contribution is simply $\nabla|L_{i,M-1}x_n|$ (resp. $\nabla|L_{M-1,j}x_n|$).

B. Experiments

In this section we consider image restoration and denoising problems in which the degradation model is given by $y = Lx + u$. In this model x, y, and u are, respectively, the original image, the recorded image, and the additive noise, while L is a known linear operator which reduces to the identity operator in denoising problems. The images have size 128×128 and are column-stacked to be represented in $\mathbb{R}^N$ ($N = 128^2$). In each experiment, the statistical hypotheses on the components of u are used to construct the closed and convex constraint set [18], [39]

$$S = \big\{z \in \mathbb{R}^N \mid \|Lz - y\|^2 \le \delta\big\}. \tag{23}$$

If L is not invertible, S is not bounded, which, strictly speaking, violates the compactness assumption of Section II-C. However, to comply with this assumption, it will suffice to replace S by $B \cap S$, where B is a large closed ball. Knowing that the original image has block features, the total variation objective is chosen as the optimality criterion. The image restoration/denoising problem then takes the form of the constrained total variation minimization program

$$\text{Find} \;\; x^* \in S \;\; \text{such that} \;\; J_{\rm tv}(x^*) = \inf J_{\rm tv}(S). \tag{24}$$

We solve this program with Algorithm 6. Let us emphasize that in the literature the standard approach to solve (24) is to modify it in order to apply a conventional algorithm e.g., [10], [29], [35], [41]. Here, there is no need to simplify, approximate, or otherwise alter (24) since Algorithm 6 can handle it as is.

Algorithm 6 is initialized with $v = 0$, $\varepsilon = 200$, and $\lambda = 0.5$. In the restoration experiment, the degraded image of Fig. 4 is obtained by convolving the original image shown in Fig. 3 with a 7×7 uniform blurring kernel and adding zero mean Gaussian white noise. The blurred image-to-noise ratio is 23.25 dB and the projector P_S is implemented by the method described in [39]. The restored image is shown in Fig. 5. In the denoising experiment, the noisy image shown in Fig. 6 is obtained by adding a zero mean Gaussian white noise to the original image shown in Fig. 3. The image-to-noise ratio is 5.65 dB and P_S is simply the projector onto a closed ball, e.g., [14]. The denoised image is shown in Fig. 7.

V. APPLICATION TO MINIMAX SET THEORETIC IMAGE RECOVERY

A. General principle

The convex feasibility approach in image recovery consists of finding an image that satisfies all the convex constraints the image to be estimated is known to possess [11], [14], [37], [39], [43]. If $(S_i)_{0\leq i\leq m}$ are the closed convex subsets of $\mathbb{R}^N$ representing these constraints, the problem is to

$$\text{Find } \; x^* \in \bigcap_{i=0}^{m} S_i. \tag{25}$$

Since the constraint sets may be constructed from inaccurate *a priori* information and uncertain measurements, the convex feasibility problem (25) may turn out to be inconsistent, i.e., $\bigcap_{i=0}^{m} S_i = \varnothing$ [13], [17], [21], [42]. It was shown in [17] that the two distinct approaches to inconsistent signal set theoretic problems of [21], [42] on the one hand, and [13] on the other hand, could be unified and extended through the single formulation

$$\text{Find } \; x^* \in S_0 \; \text{ such that } \; J_{\rm ls}(x^*) = \inf J_{\rm ls}(S_0), \quad \text{with } \; J_{\rm ls} \colon x \mapsto \frac{1}{2}\sum_{i=1}^{m} \omega_i d(x,S_i)^2, \tag{26}$$

where $\sum_{i=1}^{m} \omega_i = 1$ and $\{\omega_i\}_{1\leq i\leq m} \subset \,]0,1]$. In this formulation, S_0 represents the hard constraint for the problem, i.e., one that must imperatively be enforced, while $J_{\rm ls}$ is the *least-squares proximity function* relative to the remaining sets $(S_i)_{1\leq i\leq m}$ representing the soft constraints. A solution to (26) is therefore an image that satisfies exactly the hard constraint and that best satisfies, in a least square distance sense, the soft constraints.

In some problems, a more conservative handling of constraint inconsistency may be more appropriate. Thus, instead of minimizing the average square distance to the soft constraint sets, one may seek to minimize the worst soft constraint violation. This is tantamount to replacing (26) by

$$\text{Find } \; x^* \in S_0 \; \text{ such that } \; J_{\max}(x^*) = \inf J_{\max}(S_0), \tag{27}$$

where $J_{\max} \colon x \mapsto \max_{1\leq i\leq m} d(x,S_i)$. Henceforth, we denote by $(P_i)_{0\leq i\leq m}$ the projectors associated with $(S_i)_{0\leq i\leq m}$.

Proposition 7 *The function $J_{\max} \colon \mathbb{R}^N \to \mathbb{R}$ is convex and its subdifferential at $x \in \mathbb{R}^N \smallsetminus \bigcap_{1\leq i\leq m} S_i$ is given by*

$$\partial J_{\max}(x) = \left\{ \frac{x - \displaystyle\sum_{i\in I(x)} \mu_i P_i(x)}{\displaystyle\max_{1\leq i\leq m} d(x,S_i)} \;\middle|\; \begin{cases} \{\mu_i\}_{i\in I(x)} \subset [0,1] \\ \sum_{i\in I(x)} \mu_i = 1 \end{cases} \right\},$$

where $I(x) = \big\{1 \leq i \leq m \;\big|\; d(x,S_i) = \max_{1\leq j\leq m} d(x,S_j)\big\}$ is the set of indices of the most remote sets from x.

On the basis of the above result, the computation of $G^g_{\alpha_n}(x_n)$, as needed at Step 4 of Algorithm 6, is quite straightforward and requires only the projection of x_n onto any of the most remote sets. Indeed, take an arbitrary $i_n \in I(x_n)$. Then $J_{\max}(x_n) = d(x_n,S_{i_n})$ and it follows from Proposition 7 that we can for instance select $g_n = (x_n - P_{i_n}(x_n))/d(x_n,S_{i_n}) \in \partial J_{\max}(x_n)$ at Step 2. Consequently, Step 4 can be executed as

$$x_{n+1} = P_0\left(x_n + \left(1 - \frac{\alpha_n}{d(x_n,S_{i_n})}\right)\left(P_{i_n}(x_n) - x_n\right)\right). \tag{28}$$

B. Connections with other image recovery algorithms

We describe two instances in which we can set $\alpha_n \equiv 0$ in Algorithm 6. In view of (28), it therefore reduces to the alternating projection scheme

$$x_0 \in S_0 \; \text{ and } \; (\forall n \in \mathbb{N}) \;\; x_{n+1} = P_0\big(P_{i_n}(x_n)\big), \quad \text{where } \; d(x_n,S_{i_n}) = \max_{1\leq i\leq m} d(x_n,S_i). \tag{29}$$

B.1 Two-set inconsistent problems

Suppose that $m = 1$ in (27), i.e., there is only one soft constraint. This type of two-set inconsistent signal feasibility problem was first investigated in [21]. Here, $J_{\max} \colon x \mapsto d(x,S_1)$, which is differentiable outside S_1 (*cf.* (3)). It can then be shown that (27) is equivalent to a fixed point problem, which allows us to set $\alpha_n \equiv 0$ [17]. Thus, we recover (29) with $i_n \equiv 1$, i.e.,

$$x_0 \in S_0 \; \text{ and } \; (\forall n \in \mathbb{N}) \;\; x_{n+1} = P_0\big(P_1(x_n)\big). \tag{30}$$

This is precisely the algorithm described in [21] and further discussed in [42] to construct an image in S_0 which lies at minimum distance from the images in S_1 (see also [22, Thm. 2]).

B.2 Consistent problems

Suppose that we are dealing with a consistent set theoretic recovery problem of type (25) with m constraints. In this case, there is no need to specify a hard constraint and we put $S_0 = \mathbb{R}^N$, whence $P_0 = \mathrm{Id}$. Thus the problem reads

$$\text{Find } \; x^* \in S^* = \bigcap_{i=1}^{m} S_i. \tag{31}$$

Since it is consistent, $\alpha^* = \inf J_{\max}(\mathbb{R}^N) = 0$ and we can indeed set $\alpha_n \equiv 0$. Thus (29) becomes

$$x_0 \in \mathbb{R}^N \; \text{ and } \; (\forall n \in \mathbb{N}) \;\; x_{n+1} = P_{i_n}(x_n), \quad \text{where } \; d(x_n,S_{i_n}) = \max_{1\leq i\leq m} d(x_n,S_i). \tag{32}$$

This scheme consists of projecting the current iterate onto one of the most distant sets from it. The convergence of (32) to a point in S^* was established in [8, Thm. 2]. In the same paper [8, Thm. 1], Brègman also proved the convergence to a point in S^* of the periodic projection scheme

$$x_0 \in \mathbb{R}^N \; \text{ and } \; (\forall n \in \mathbb{N}) \;\; x_{n+1} = P_{i_n}(x_n), \quad \text{where } \; i_n = n(\text{modulo } m) + 1, \tag{33}$$

which was popularized by [43] and became known as POCS in the signal processing community.

C. Experiment

As in Section IV, all images have size 128×128 and are column-stacked to be represented in $\mathbb{R}^N$ ($N = 128^2$). The original image x of Fig. 8 is degraded by convolutional blur with a uniform 7×7 point spread function b and addition of noise. The noise samples are distributed in the interval $[0, R]$ ($R = 5$) and the resulting blurred image-to-noise ratio is 33 dB. The degraded image y is shown in Fig. 9. It can be written as $y = Lx + u$, where L is the $N \times N$ block-Tœplitz matrix associated with b [5] and u is a noise vector.

Let us now construct the constraint sets for this problem. First, the property that the pixel values are nonnegative and do not exceed $\chi = 255$ yields the hard constraint for the problem. The associated set is the compact convex set $S_0 = [0, \chi]^N$. Next, we assume that the point spread function matrix L is known but that the range of the noise samples is incorrectly estimated to be $[\beta_1, \beta_2]$, where $\beta_1 = 3/2$ and $\beta_2 = 7/2$. This information leads to the N hyperslabs

$$S_{p+1} = \big\{ z \in \mathbb{R}^N \mid \beta_1 \leq y^p - \langle L_p \mid z \rangle \leq \beta_2 \big\} \tag{34}$$

where y^p is the pth component of y and L_p is the pth row of L for $0 \leq p \leq N-1$ [18]. Finally, it is assumed that the discrete Fourier transform (DFT), $\widehat{x}$, of x is known over the low frequency range

$$K' = \big\{ (k,l) \in \{0, \ldots, 127\}^2 \mid 0 \leq k, l \leq F \big\}, \text{ where } F = 3. \tag{35}$$

Recall that the DFT of the real image x possesses the conjugate-symmetry properties

$$\begin{cases} \widehat{x}(k,0) = \overline{\widehat{x}}(F-k,0) & \text{if } k \neq 0 \\ \widehat{x}(0,l) = \overline{\widehat{x}}(0,F-l) & \text{if } l \neq 0 \\ \widehat{x}(k,l) = \overline{\widehat{x}}(F-k,F-l) & \text{if } kl \neq 0 \end{cases} \tag{36}$$

for all $(k,l) \in \{0, \cdots, 127\}^2$. Accordingly, the set K' must be extended to a set K including all the symmetric pairs. Thus, $S_m = \big\{ z \in \mathbb{R}^N \mid \widehat{z} 1_K = \widehat{x} 1_K \big\}$, where $m = N+1$ and 1_K is the characteristic function of the set K, which takes value 1 on K and 0 on its complement. The projectors onto the sets $(S_i)_{0 \leq i \leq m}$ are straightforward and can be found in [14].

The above set theoretic formulation is rendered inconsistent by the fact that the bounds on the amplitude of the noise are incorrect. The problem is set up as (27) and solved by Algorithm 6, where $v = 0$, $\varepsilon = 10^{-3}$, and $\lambda = 0.5$. The restored image is shown in Fig. 10.

VI. Conclusion

The use of nondifferentiable objectives has been advocated in various image recovery studies. In this paper, we have proposed a reliable, general-purpose algorithm for recovering an image by minimizing a nondifferentiable convex function over a convex feasibility set. Its principle is to alternate a subgradient projection onto an adaptively refined approximation to the optimal level set of the objective and an exact projection onto the feasibility set. Unlike the methods typically in use in image recovery, the proposed algorithm is not tailored to a specific kind of nondifferentiable objective and does not require any alteration of the problem formulation. Numerical applications to image denoising and restoration problems have confirmed the efficiency of the method. For problems in which the feasibility set is complex, eliminating the reliance on projections to enforce feasibility would further enhance the efficiency of the method and constitutes a high priority for further work.

Appendix A – Proofs

Proof of Theorem 4: In view of (17), the ability to detect infeasibility at Step 5 implies that Step 6 will be executed repeatedly until we get $\eta_{n+1} \leq \lambda\varepsilon$ at Step 1 and terminate the algorithm at iteration $n+1$ with x_n as a solution. Note that $\eta_n = \eta_{n+1}/\lambda \leq \varepsilon$ and that, since infeasibility has been detected at iteration n, $\alpha_n < \alpha^*$. Consequently, it follows from (9) and (11) that $J(x_n) \leq \overline{\alpha}_n = \alpha_n + \eta_n \leq \alpha^* + \varepsilon$. ∎

Proof of Proposition 5: (i): Fix $m \in \mathbb{N}$. It follows from (11) and the definition of $\mathbb{N}_m$ that, for every $n \in \mathbb{N}_m$, $\alpha_n = \overline{\alpha}_n - \lambda^m \eta_0$. Hence, since $(\overline{\alpha}_n)_{n \in \mathbb{N}_m}$ is nonincreasing by (10), so is $(\alpha_n)_{n \in \mathbb{N}_m}$. To prove the second claim, we proceed by contradiction. Suppose that, for some $m \in \mathbb{N}$, we have $(\forall n \in \mathbb{N}_m)$ $\alpha_n \geq \alpha^*$. Then $(\forall n \in \mathbb{N}_m)$ $S \cap \operatorname{lev}_{\leq \alpha_n} J \neq \varnothing$ and therefore $\mathbb{N}_m$ is infinite by (11). Since $(\alpha_n)_{n \in \mathbb{N}_m}$ is nonincreasing and bounded from below by α^*, it converges to some $\alpha \geq \alpha^*$. It follows from Theorem 1 that $(x_n)_{n \in \mathbb{N}_m}$ converges to some $x \in S \cap \operatorname{lev}_{\leq \alpha} J$ and, from the continuity of J, that $(J(x_n))_{n \in \mathbb{N}_m}$ converges to $J(x) \leq \alpha$. On the other hand, (11) and (9) yield

$$(\forall n \in \mathbb{N}_m) \;\; \alpha_n = \overline{\alpha}_n - \lambda^m \eta_0 \leq J(x_n) - \lambda^m \eta_0. \tag{A1}$$

Hence, by passing to the limit, we obtain $\alpha \leq J(x) - \lambda^m \eta_0$. Altogether, $\alpha < J(x) \leq \alpha$, which is absurd.

(ii): Suppose that $\mathbb{N}_m$ is infinite. Then $(\forall n \in \mathbb{N}_m)$ $\alpha_n = \overline{\alpha}_n - \lambda^m \eta_0 \geq \alpha^* - \lambda^m \eta_0$ and it follows from (i) that $(\alpha_n)_{n \in \mathbb{N}_m}$ converges to some $\alpha \in \left]-\infty, \alpha^*\right[$. Since $(x_n)_{n \in \mathbb{N}_m}$ lies in the compact set S, it contains a subsequence $(x_{k_n})_{n \geq 0}$ converging to some point $x \in S$ and it follows from the continuity of J that

$$\lim J(x_{k_n}) - \alpha_{k_n} = J(x) - \alpha. \tag{A2}$$

Let us now show that $\sum_{n \in \mathbb{N}_m} \rho_n = +\infty$. Given $\beta \in \mathbb{R}$, we shall use the notation $\beta^+ = \max\{0, \beta\}$. First, let us note that, since $(x_{k_n})_{n \geq 0}$ is bounded, [34, Thm. 24.7] implies that

$$\sup_{n \geq 0} \|g(x_{k_n})\| < +\infty. \tag{A3}$$

Hence, $\sum_{n \in \mathbb{N}_m} \rho_n < +\infty \Rightarrow \rho_{k_n} \to 0 \Rightarrow$ [by (15)] $G^g_{\alpha_{k_n}}(x_{k_n}) - x_{k_n} \to 0 \Rightarrow$ [by (4)] $\big(J(x_{k_n}) - \alpha_{k_n}\big)^+ / \|g(x_{k_n})\| \to 0 \Rightarrow$ [by (A3)] $\big(J(x_{k_n}) - \alpha_{k_n}\big)^+ \to 0 \Rightarrow$ [continuity of $\beta \mapsto \beta^+$] $\big(\lim J(x_{k_n}) - \alpha_{k_n}\big)^+ = 0 \Rightarrow$ [by (A2)] $\big(J(x) - \alpha\big)^+ = 0 \Rightarrow J(x) \leq \alpha$. However this is absurd since $x \in S \Rightarrow J(x) \geq \alpha^* > \alpha$.

(iii): Fix $n \in \mathbb{N}_m$, $k \in \{l_m, \ldots, n\}$, and set $x^* = P_{S^*}(x_{l_m})$. If $S \cap \operatorname{lev}_{\leq \alpha_k} J \neq \varnothing$, then $x^* \in S^* = S \cap \operatorname{lev}_{\leq \alpha^*} J \subset S \cap \operatorname{lev}_{\leq \alpha_k} J$. It then follows from (5) and (2) that

$$\begin{aligned} \rho_k &= \|G^g_{\alpha_k}(x_k) - x_k\|^2 + \|P_S \circ G^g_{\alpha_k}(x_k) - G^g_{\alpha_k}(x_k)\|^2 \\ &\leq \|x_k - x^*\|^2 - \|G^g_{\alpha_k}(x_k) - x^*\|^2 + \|G^g_{\alpha_k}(x_k) - x^*\|^2 \\ &\quad - \|P_S \circ G^g_{\alpha_k}(x_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2. \end{aligned} \tag{A4}$$

Now suppose $S \cap \operatorname{lev}_{\leq \alpha_n} J \neq \varnothing$. Then $S \cap \bigcap_{k=l_m}^{n} \operatorname{lev}_{\leq \alpha_k} J \neq \varnothing$ and (A4) implies

$$\sum_{k=l_m}^{n} \rho_k \leq \|x_{l_m} - x^*\|^2 - \|x_{n+1} - x^*\|^2. \tag{A5}$$

Consequently,

$$\begin{aligned} &\|x_{l_m} - x^*\|^2 - \|x_{n+1} - x^*\|^2 \\ &\quad = \; 2\langle x_{l_m} - x^* \mid x_{l_m} - x_{n+1}\rangle - \|x_{l_m} - x_{n+1}\|^2 \\ &\quad \leq \; 2\|x_{l_m} - x^*\| \cdot \|x_{l_m} - x_{n+1}\| - \|x_{l_m} - x_{n+1}\|^2 \\ &\quad \leq \; 2\gamma_m \|x_{l_m} - x_{n+1}\| - \|x_{l_m} - x_{n+1}\|^2. \end{aligned} \tag{A6}$$

Thus, $S \cap \operatorname{lev}_{\leq \alpha_n} J \neq \varnothing$ implies

$$\sum_{k=l_m}^{n} \rho_k \leq \|x_{l_m} - x_{n+1}\|(2\gamma_m - \|x_{l_m} - x_{n+1}\|). \tag{A7}$$

■

Proof of Proposition 7: Let $\operatorname{conv} Q$ be the convex hull of a set Q in $\mathbb{R}^N$, i.e., the smallest convex set containing Q, and let $(J_i)_{1 \leq i \leq m}$ be convex functions from $\mathbb{R}^N$ into $\mathbb{R}$. Then by [24, Thm. VI.4.4.2] $J = \max_{1 \leq i \leq m} J_i$ is convex and, for every $x \in \mathbb{R}^N$, $\partial J(x) = \operatorname{conv} \bigcup_{i \in I(x)} \partial J_i(x)$, where $I(x) = \big\{1 \leq i \leq m \mid J_i(x) = J(x)\big\}$. Hence the claim follows from (3). ■

References

[1] R. Acar and C. R. Vogel, "Analysis of bounded variation penalty methods for ill-posed problems," *Inverse Problems,* vol. 10, pp. 1217-1229, 1994.

[2] E. Allen, R. Helgason, J. Kennington, and B. Shetty, "A generalization of Polyak's convergence result for subgradient optimization," *Math. Programming,* vol. 37, pp. 309-317, 1987.

[3] S. Alliney, "A property of the minimum vectors of a regularizing functional defined by means of the absolute norm," *IEEE Trans. Signal Process.,* vol. 45, pp. 913-917, Apr. 1997.

[4] S. Alliney and S. A. Ruzinsky, "An algorithm for the minimization of mixed ℓ_1 and ℓ_2 norms with application to Bayesian estimation," *IEEE Trans. Signal Process.,* vol. 42, pp. 618-627, Mar. 1994.

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

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

[7] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Belmont, MA: Athena Scientific, 1997.

[8] L. M. Brègman, "The method of successive projection for finding a common point of convex sets," *Soviet Math. Dokl.,* vol. 6, pp. 688-692, May 1965.

[9] Y. Censor and G. T. Herman, "On some optimization techniques in image reconstruction from projections," *Appl. Numer. Math.,* vol. 3, pp. 365-391, 1987.

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

[11] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE,* vol. 81, pp. 182-208, Feb. 1993.

[12] P. L. Combettes, "Signal recovery by best feasible approximation," *IEEE Trans. Image Process.,* vol. 2, pp. 269-271, Apr. 1993.

[13] P. L. Combettes, "Inconsistent signal feasibility problems: Least-squares solutions in a product space," *IEEE Trans. Signal Process.,* vol. 42, pp. 2955-2966, Nov. 1994.

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

[15] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Trans. Image Process.,* vol. 6, pp. 493-506, Apr. 1997.

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

[17] P. L. Combettes and P. Bondon, "Hard-constrained inconsistent signal feasibility problems," *IEEE Trans. Signal Process.,* vol. 47, pp. 2460-2468, Sept. 1999.

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

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

[20] J.-L. Goffin and K. C. Kiwiel, "Convergence of a simple subgradient level method," *Math. Programming,* vol. 85, pp. 207-211, 1999.

[21] M. Goldburg and R. J. Marks II, "Signal synthesis in the presence of an inconsistent set of constraints," *IEEE Trans. Circuits Syst.,* vol. 32, pp. 647-663, July 1985.

[22] L. G. Gubin, B. T. Polyak, and E. V. Raik, "The method of projections for finding the common point of convex sets," *USSR Comput. Math. Math. Phys.,* vol. 7, pp. 1-24, 1967.

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

[24] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms.* New York: Springer-Verlag, 1993.

[25] S. Kim, H. Ahn, and S. C. Cho, "Variable target value subgradient method," *Math. Programming,* vol. 49, pp. 359-369, 1991.

[26] K. C. Kiwiel, "The efficiency of subgradient projection methods for convex optimization. Part I: General level methods," and "Part II: Implementations and extensions," *SIAM J. Control Optim.,* vol. 34, pp. 660-676 and 677-697, Mar. 1996.

[27] R. M. Leahy and C. E. Goutis, "An optimal technique for constraint-based image restoration and reconstruction," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 34, pp. 1629-1642, Dec. 1986.

[28] C. Lemaréchal, "Nondifferentiable optimization," in *Handbooks in Operations Research and Management Science,* vol. 1, pp. 529-572. New York: North-Holland, 1989.

[29] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image reconstruction," *IEEE Trans. Image Process.,* vol. 5, pp. 987-995, June 1996.

[30] D. R. Luke, J. V. Burke, and R. G. Lyon, "Optical wavefront reconstruction: Theory and numerical methods," *SIAM Rev.,* vol. 44, pp. 169-224, 2002.

[31] D. Noll, "Reconstruction with noisy data: An approach via eigenvalue optimization," *SIAM J. Optim.,* vol. 8, pp. 82-104, Feb. 1998.

[32] B. T. Polyak, "Minimization of unsmooth functionals," *USSR Comput. Math. Math. Phys.,* vol. 9, pp. 14-29, 1969.

[33] B. T. Polyak, *Introduction to Optimization.* New York: Optimization Software, 1987.

[34] R. T. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton University Press, 1970.

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

[36] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions.* New York: Springer-Verlag, 1985.

[37] H. Stark (Editor), *Image Recovery: Theory and Application.* San Diego, CA: Academic Press, 1987.

[38] D. M. Titterington, "General structure of regularization procedures in image reconstruction," *Astron. Astrophys.,* vol. 144, pp. 381-387, Mar. 1985.

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

[40] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.,* vol. 17, pp. 227-238, Jan. 1996.

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

[42] D. C. Youla and V. Velasco, "Extensions of a result on the synthesis of signals in the presence of inconsistent constraints," *IEEE Trans. Circuits Syst.,* vol. 33, pp. 465-468, Apr. 1986.

[43] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 – theory," *IEEE Trans. Medical Imaging,* vol. 1, pp. 81-94, Oct. 1982.

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

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

Jian Luo was born in Inner Mongolia, China, on February 5, 1964. He received the B.E. degree in electrical engineering from the Taiyuan Heavy Machinery Institute, in 1986. He received the Ph.D. degree in electrical engineering from the City University of New York, New York, NY, in 2000.

From 1986 to 1992 he was an assistant electrical engineer with the Chengdu Tool Research Institute of the Ministry of Machine Building and Electronics Industry of China. From 2000 to 2002 he was a development engineer at Gentner Communications Corporation.

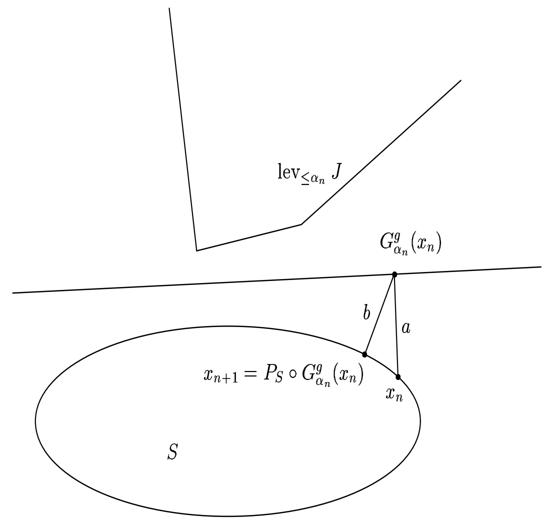

Fig. 1. $G^g_{\alpha_n}(x_n)$ is a subgradient projection of x_n onto $\mathrm{lev}_{\alpha_n} J$ and $\rho_n = a^2 + b^2$ (here $\alpha_n < \alpha^*$).

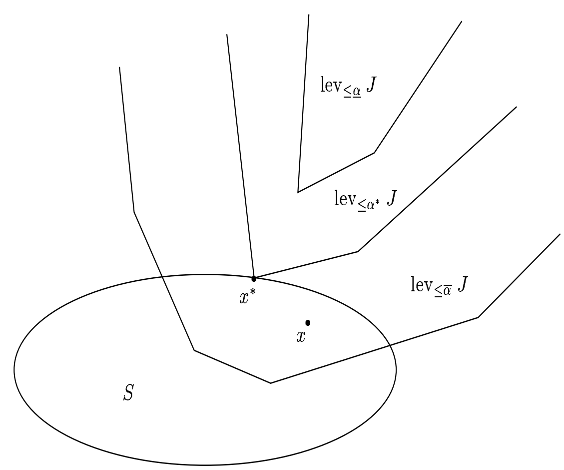

Fig. 2. Level sets of the function J for $\underline{\alpha} \leq \alpha^* \leq \overline{\alpha}$. x^* is an optimal solution and x a point in $S \cap \mathrm{lev}_{\leq \overline{\alpha}} J$.

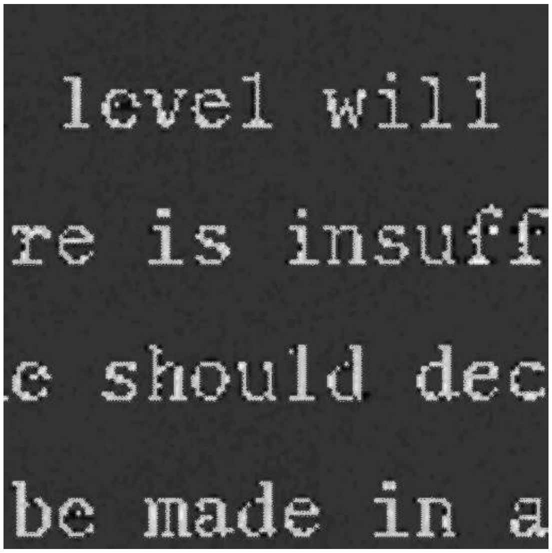

Fig. 3. Original image.

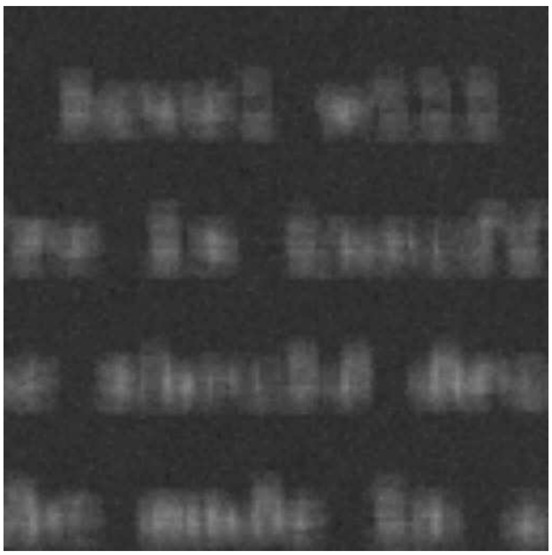

Fig. 4. Degraded image.

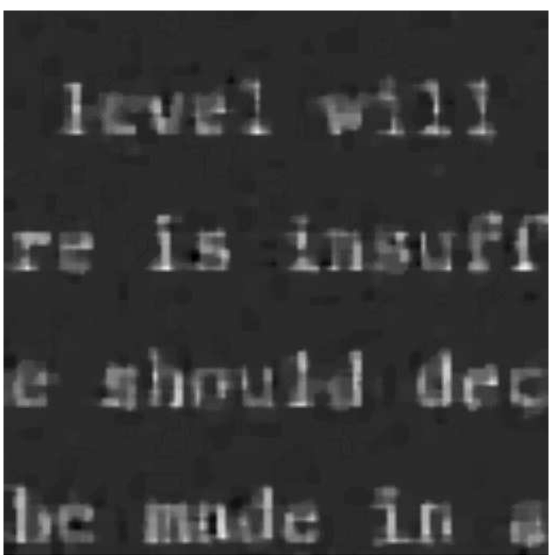

Fig. 5. Restored image.

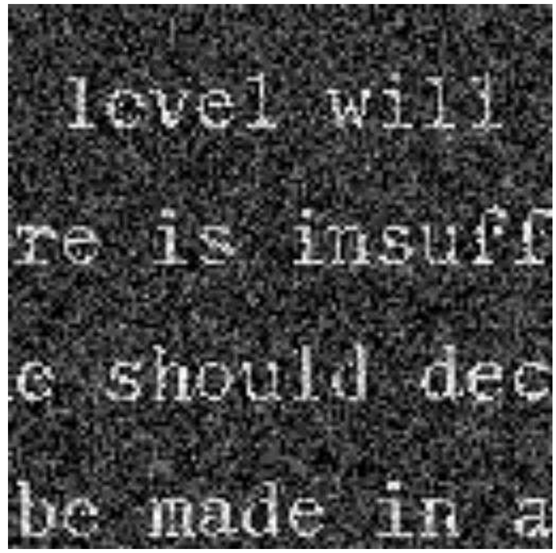

Fig. 6. Noisy image.

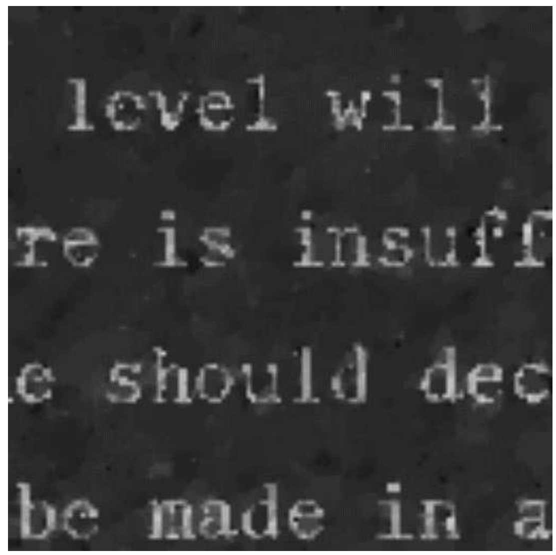

Fig. 7. Denoised image.

Fig. 8. Original image.

Fig. 9. Degraded image.

Fig. 10. Restored image.