# FULLY PROXIMAL SPLITTING ALGORITHMS IN IMAGE RECOVERY

1st Patrick L. Combettes
*Department of Mathematics*
*North Carolina State University*
Raleigh, NC 27695-8205, USA

2nd Lilian E. Glaudin
*Laboratoire Jacques-Louis Lions*
*Sorbonne Université*
Paris, France

***Abstract*—Structured convex optimization problems in image recovery typically involve a mix of smooth and nonsmooth functions. The common practice is to activate the smooth functions via their gradient and the nonsmooth ones via their proximity operator. We show that, although intuitively natural, this approach is not necessarily the most efficient numerically and that, in particular, activating all the functions proximally may be advantageous. To make this viewpoint viable computationally, we derive a number of new examples of proximity operators of smooth convex functions arising in applications.**

***Index Terms*—convex optimization, image recovery, nonsmooth optimization, proximal splitting algorithm, proximity operator**

## I. Introduction

Splitting in convex optimization methods for image recovery can be traced back to the influential work of Youla [18], [19]. The convex feasibility framework he proposed consists in formulating the image recovery problem as that of finding an image in a Euclidean space $\mathcal{H}$ satisfying $m$ constraints derived from *a priori* knowledge and the observed data. The constraints are represented by closed convex sets $(C_i)_{1\leqslant i\leqslant m}$ and the problem is therefore to

$$\text{find } x \in \bigcap_{i=1}^{m} C_i. \tag{1}$$

Now, for every $i \in \{1,\dots,m\}$, let $\text{proj}_{C_i}$ be the projection operator onto $C_i$, which maps each point $x \in \mathcal{H}$ to its unique closest point in $C_i$. The methodology of projection methods is to split the problem of finding a point in $\bigcap_{i=1}^{m} C_i$ into a sequence of simpler problems involving the sets $(C_i)_{1\leqslant i\leqslant m}$ individually [10]. For instance, the POCS (Projection Onto Convex Sets) algorithm advocated in [19] is governed by the updating rule

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (\text{proj}_{C_1} \circ \cdots \circ \text{proj}_{C_m})x_n. \tag{2}$$

Projection operators are of limited use beyond feasibility and best approximation problems. Modern image recovery convex variational formulations have a complex structure that requires sophisticated analysis tools and solution methods. To solve such formulations while preserving the spirit of splitting methods in feasibility problems, one approach is to use an extended notion of a projection operator. In [14] it was suggested to use Moreau's notion of a proximity operator [15] for this purpose. Recall that the proximity operator of a proper lower semicontinuous convex function $\varphi\colon \mathcal{H} \to ]-\infty,+\infty]$ is

$$\text{prox}_{\varphi}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \underset{y\in\mathcal{H}}{\text{argmin}} \left(\varphi(y) + \frac{1}{2}\|x-y\|^2\right), \tag{3}$$

and that it reduces to $\text{proj}_{C_i}$ when $\varphi = \iota_{C_i}$, the indicator function of $C_i$. We refer the reader to [3, Chapter 24] for a detailed account of the properties of proximity operators with various examples, to [12] for a tutorial on proximal methods in signal processing, and to [2], [7], [9], [16], [17] for specific applications to image recovery.

A prevalent viewpoint in modern proximal splitting algorithms is that to activate each function $\varphi$ appearing in the model there are two options: if $\varphi$ is smooth, i.e., real-valued and differentiable everywhere with a Lipschitzian gradient, then use $\nabla\varphi$; otherwise, use $\varphi$ proximally, i.e., via its proximity operator (3).

We propose a more nuanced viewpoint and submit that, when $\varphi$ is smooth, it may be computationally advantageous to activate it proximally when its proximity operator can be implemented. To motivate this viewpoint, let us first observe that a tight Lipschitz constant for the gradient of $\varphi$ may not be easy to estimate (see, e.g., [1], [4], [5]), which limits the range of the proximal parameters and may have a detrimental incidence on the speed of convergence. Our second observation is that proximal steps behave numerically quite differently from gradient steps, which may have a positive impact on the asymptotic behavior of the algorithm (see Fig. 1).

The paper is organized as follows. Section II provides explicit expressions for proximity operators of smooth convex functions commonly encountered in image recovery. Image recovery applications are presented in Sections III and IV. Numerical comparisons between the forward-backward method and the Douglas-Rachford method are presented in Section III. Next, numerical comparisons between splitting algorithms in which smooth

The work of P. L. Combettes was supported by the National Science Foundation under grant CCF-1715671.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

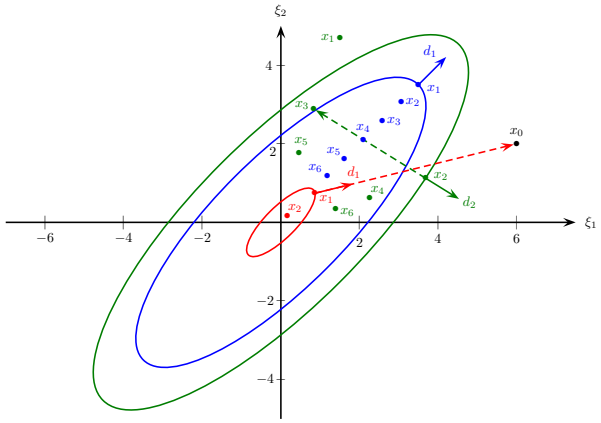


Fig. 1. Comparison of the gradient method $x_{n+1} = x_n - \gamma\nabla\varphi(x_n)$ (in green) and of the proximal point algorithm $x_{n+1} = \operatorname{prox}_{\gamma\varphi} x_n$ (in red) in $\mathcal{H} = \mathbb{R}^2$ for $\varphi\colon (\xi_1, \xi_2) \mapsto 9\xi_1^2 - 14\xi_1\xi_2 + 9\xi_2^2$. The gradient method is implemented with $\gamma = 1.8/\beta$ as this choice gave rise to the fastest convergence. On the other hand, the proximal point algorithm is implemented with the default choice $\gamma = 1$ (larger values gave even faster convergence). The two algorithms behave quite differently, both in terms of directions of movement and of trajectories. At iteration $n$, call $d_n = \nabla\varphi(x_n)/\|\nabla\varphi(x_n)\|$ the normalized gradient at $x_n$. Consider the action of the gradient, say at iteration $n = 2$. The next iterate $x_3$ is obtained by moving from $x_2$ in the direction opposite to the gradient at $x_2$. By contrast, consider the action of the proximal point algorithm, say at iteration $n = 0$. The next iterate $x_1$ satisfies the implicit equation $x_0 - x_1 = \gamma\nabla\varphi(x_1)$, which means that $x_1$ is obtained by moving from $x_0$ in the direction opposite to the gradient at $x_1$. Finally, we include the orbit (in blue) of the inertial version of the gradient method obtained by setting $f = 0$ and $h = \varphi$ in algorithm (5), and choosing the parameters $\alpha = 2.01$ and $\gamma = 1/\beta$, which gave the fastest convergence.

functions are activated via gradient steps and those in which all functions are activated via their proximity operators are conducted. More extensive experiments on various problems with additional algorithms can be found in [11]. While no universal conclusion may be drawn from these experiments, they suggest that fully proximal splitting algorithms deserve to be given serious consideration in applications.

**Notation.** The notation follows that of [3]. Throughout, $\mathcal{H}$, $\mathcal{G}$, and $\mathcal{G}_k$ are real Euclidean spaces and $\mathcal{B}(\mathcal{H}, \mathcal{G})$ is the space of linear operators from $\mathcal{H}$ to $\mathcal{G}$.

## II. Proximity operators of smooth convex functions

We show that a broad range of smooth convex functions encountered in signal recovery actually have an explicit proximity operator. The proof of the following results will be found in [11].

**Example II.1** Let $I$ be a nonempty finite set. For every $i \in I$, let $\mathcal{G}_i$ be a real Hilbert space, let $V_i$ be a closed vector subspace of $\mathcal{G}_i$, let $r_i \in \mathcal{G}_i$, let $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{G}_i)$, and let $\alpha_i \in ]0, +\infty[$. Set $h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\sum_{i\in I} \alpha_i d_{V_i}^2(L_i x - r_i)$ and $Q = (\mathrm{Id} + \gamma\sum_{i\in I} \alpha_i L_i^* \operatorname{proj}_{V_i^\perp} L_i)^{-1}$. Let $\gamma \in ]0, +\infty[$, set $\beta = \sum_{i\in I} \alpha_i \|L_i\|^2$, and let $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and differentiable with a $\beta$-Lipschitzian gradient, $\nabla h(x) = \sum_{i\in I} \alpha_i L_i^* \operatorname{proj}_{V_i^\perp}(L_i x - r_i)$, and $\operatorname{prox}_{\gamma h} x = Q(x + \gamma\sum_{i\in I} \alpha_i L_i^* \operatorname{proj}_{V_i^\perp} r_i)$.

In many digital image recovery applications, the operators $(L_i)_{i\in I}$ are representable by block-circulant matrices and the computation of $Q$ in Example II.1 is therefore straightforward. If not, note that $Q$ is computable efficiently as the inverse of a positive definite symmetric matrix; by contrast $\beta$ is much more expensive to compute. The next construction, which involves the distance function $d_C$ to a convex set $C$, captures a broad range of functions of interest.

**Example II.2** Let $C$ be a nonempty closed convex subset of $\mathcal{H}$, let $\beta \in ]0, +\infty[$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a $\beta$-Lipschitzian derivative, and set $h = \phi \circ d_C$. Let $\gamma \in ]0, +\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and differentiable with a $\beta$-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\phi'(d_C(x))}{d_C(x)}(x - \operatorname{proj}_C x), & \text{if } x \notin C; \\ 0, & \text{if } x \in C, \end{cases}$$

and $\operatorname{prox}_{\gamma h} x =$

$$\begin{cases} \operatorname{proj}_C x + \dfrac{\operatorname{prox}_{\gamma\phi} d_C(x)}{d_C(x)}(x - \operatorname{proj}_C x), & \text{if } x \notin C; \\ x, & \text{if } x \in C. \end{cases}$$

**Example II.3 (abstract smooth Vapnik loss function)** Let $\varepsilon \in ]0, +\infty[$, let $\beta \in ]0, +\infty[$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a $\beta$-Lipschitzian derivative, let $\vartheta\colon \xi \mapsto \max(|\cdot| - \varepsilon, 0)$ be the standard Vapnik loss function, and set $h = \phi \circ \vartheta \circ \|\cdot\|$. Let $\gamma \in ]0, +\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and differentiable with a $\beta$-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\phi'(\|x\| - \varepsilon)}{\|x\|} x, & \text{if } \|x\| > \varepsilon; \\ 0, & \text{if } \|x\| \leqslant \varepsilon, \end{cases}$$

and

$$\operatorname{prox}_{\gamma h} x = \begin{cases} \dfrac{\varepsilon + \operatorname{prox}_{\gamma\phi}(\|x\| - \varepsilon)}{\|x\|} x, & \text{if } \|x\| > \varepsilon; \\ x, & \text{if } \|x\| \leqslant \varepsilon. \end{cases}$$

**Example II.4 (abstract Huber function)** Let $C$ be a nonempty closed convex subset of $\mathcal{H}$, let $\rho \in ]0, +\infty[$, and set

$$h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \begin{cases} \rho d_C(x) - \rho^2/2, & \text{if } d_C(x) > \rho; \\ d_C(x)^2/2, & \text{if } d_C(x) \leqslant \rho. \end{cases} \tag{4}$$

Let $\gamma \in ]0, +\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and differentiable with a nonexpansive gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\rho}{d_C(x)}(x - \operatorname{proj}_C x), & \text{if } d_C(x) > \rho; \\ x - \operatorname{proj}_C x, & \text{if } d_C(x) \leqslant \rho, \end{cases}$$

and $\operatorname{prox}_{\gamma h} x =$

$$\begin{cases} x + \dfrac{\gamma\rho}{d_C(x)}(\operatorname{proj}_C x - x), & \text{if } d_C(x) > (\gamma+1)\rho; \\ \dfrac{1}{\gamma+1}\big(x + \gamma \operatorname{proj}_C x\big), & \text{if } d_C(x) \leqslant (\gamma+1)\rho. \end{cases}$$

The following extension of Example II.2 involves a composition with a linear operator.

**Example II.5** Let $M \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that $MM^* = \theta \operatorname{Id}$ for some $\theta \in ]0, +\infty[$. Let $D$ be a nonempty closed convex subset of $\mathcal{G}$, let $\mu \in ]0, +\infty[$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a $\mu$-Lipschitzian derivative, and set $h = \phi \circ d_D \circ M$ and $\beta = \mu\|M\|^2$. Let $\gamma \in ]0, +\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and differentiable with a $\beta$-Lipschitzian gradient, and $\operatorname{prox}_{\gamma h} x =$

$$\begin{cases} x + \dfrac{\theta^{-1}\big(d_D(Mx) - \operatorname{prox}_{\gamma\theta\phi} d_D(Mx)\big)}{d_D(Mx)} \\ \qquad\qquad \times M^*\big(\operatorname{proj}_D(Mx) - Mx\big), & \text{if } Mx \notin D; \\ x, & \text{if } Mx \in D. \end{cases}$$

The condition $MM^* = \theta \operatorname{Id}$ used in Example II.5 arises in particular in problems involving tight frame representations [8]. When it is not satisfied, one can still deal with smooth functions of the type $\phi \circ d_C \circ M$ in modern structured proximal splitting techniques by activating $\operatorname{prox}_{\phi \circ d_C}$ and $M$ separately.

**Example II.6** Suppose that $\mathcal{H}$ is separable and that $\varnothing \neq \mathbb{K} \subset \mathbb{N}$, and let $(e_k)_{k\in\mathbb{K}}$ be an orthonormal basis of $\mathcal{H}$. For every $k \in \mathbb{K}$, let $\beta_k \in ]0, +\infty[$ and let $\phi_k\colon \mathbb{R} \to \mathbb{R}$ be a differentiable convex function such that $\phi_k \geqslant \phi_k(0) = 0$ and $\phi_k'$ is $\beta_k$-Lipschitzian. Suppose that $\beta = \sup_{k\in\mathbb{K}} \beta_k < +\infty$ and define $(\forall x \in \mathcal{H})\ h(x) = \sum_{k\in\mathbb{K}} \phi_k(\langle x \mid e_k\rangle)$. Let $\gamma \in ]0, +\infty[$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and differentiable with a $\beta$-Lipschitzian gradient,

$$(\forall x \in \mathcal{H}) \quad \nabla h(x) = \sum_{k\in\mathbb{K}} \phi_k'(\langle x \mid e_k\rangle) e_k,$$

and $(\forall x \in \mathcal{H})\ \operatorname{prox}_{\gamma h} x = \sum_{k\in\mathbb{K}} \big(\operatorname{prox}_{\gamma\phi_k}\langle x \mid e_k\rangle\big) e_k$.

## III. Forward-Backward versus Douglas-Rachford splitting

We compare the numerical behavior of the forward-backward algorithm (we choose the inertial implementation of [6]) with that of the Douglas-Rachford algorithm, which is a fully proximal method.

**Proposition III.1 (inertial forward-backward)** *Let $\beta \in ]0, +\infty[$, let $f \in \Gamma_0(\mathcal{H})$, let $h\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a $\beta$-Lipschitzian gradient, let $\gamma \in ]0, 1/\beta]$, let $\alpha \in ]2, +\infty[$, and suppose that $\operatorname{Argmin}(f + h) \neq \varnothing$. Let $x_0 = x_{-1} \in \mathcal{H}$ and iterate for $n = 0, 1, \ldots$*

$$\left\lfloor \begin{aligned} z_n &= x_n + \frac{n-1}{n+\alpha}(x_n - x_{n-1}) \\ y_n &= z_n - \gamma\nabla h(z_n) \\ x_{n+1} &= \operatorname{prox}_{\gamma f} y_n. \end{aligned} \right. \tag{5}$$

*Then there exists $x \in \operatorname{Argmin}(f + h)$ such that $x_n \to x$.*

**Proposition III.2 (Douglas-Rachford)** [3, Cor. 28.3] *Let $f$ and $g$ be functions in $\Gamma_0(\mathcal{H})$ such that $\operatorname{Argmin}(f+g) \neq \varnothing$ and the relative interiors of $\operatorname{dom} f$ and $\operatorname{dom} g$ intersect, let $\gamma \in ]0, +\infty[$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$. Let $y_0 \in \mathcal{H}$ and iterate for $n = 0, 1, \ldots$*

$$\left\lfloor \begin{aligned} z_n &= \operatorname{prox}_{\gamma g} y_n \\ x_n &= \operatorname{prox}_{\gamma f}(2z_n - y_n) \\ y_{n+1} &= y_n + \lambda_n(x_n - z_n). \end{aligned} \right. \tag{6}$$

*Then there exists $x \in \operatorname{Argmin}(f + g)$ such that $x_n \to x$.*

We consider a digital image restoration problem. All images have $128 \times 128$ pixels and therefore the underlying Euclidean space is $\mathcal{H} = \mathbb{R}^N$ ($N = 128^2$) equipped with the standard Euclidean norm $\|\cdot\|_2$. We have run many instances of each problem under various configurations, and the limited numerical results we present here are representative of the behavior one can expect. Note that the results are given in terms of iteration numbers as the algorithms compared in each experiment have similar execution time per iteration.

The original image is $\overline{x}$ and the degraded image is $y = H\overline{x} + w$, where $H$ models a convolution with a uniform rectangular kernel of size $15 \times 5$ and $w$ is a Gaussian white noise realization (see Fig. 2(a)–(b)). The blurred image-to-noise-ratio is 15.5 dB. Since each pixel value is known to be in $[0, 255]$, we use the hard constraint set $C = [0, 255]^N$. As is customary, the natural sparsity of $\overline{x}$ is promoted using the function $\|\cdot\|_1$. Altogether, the problem is

$$\min_{x\in C} \ \|x\|_1 + \frac{1}{2}\|Hx - y\|_2^2. \tag{7}$$

Now set $f = \|\cdot\|_1 + \iota_C$ and $h = \|H \cdot - y\|_2^2/2$, and let $\gamma \in ]0, +\infty[$. Then $f \in \Gamma_0(\mathcal{H})$ has an explicit proximity operator $\operatorname{prox}_{\gamma f} = \operatorname{proj}_C \circ \operatorname{soft}_\gamma$, where $\operatorname{soft}_\gamma$ is the soft thresholder on $[-\gamma, \gamma]$. Furthermore, $h$ is smooth and its proximity operator is provided by Example II.1. We solve (7) both with (5) and (6). The algorithms have similar iteration complexity. The algorithms are initialized at zero and implemented with parameters for which they perform better, that is, $\gamma = 1/\beta$ and $\alpha = 3$ for (5), and $\gamma = 30$ and $\lambda_n \equiv 1.9$ for (6). The results of Figs. 2–3 show a superior performance for the fully proximal algorithm (6).

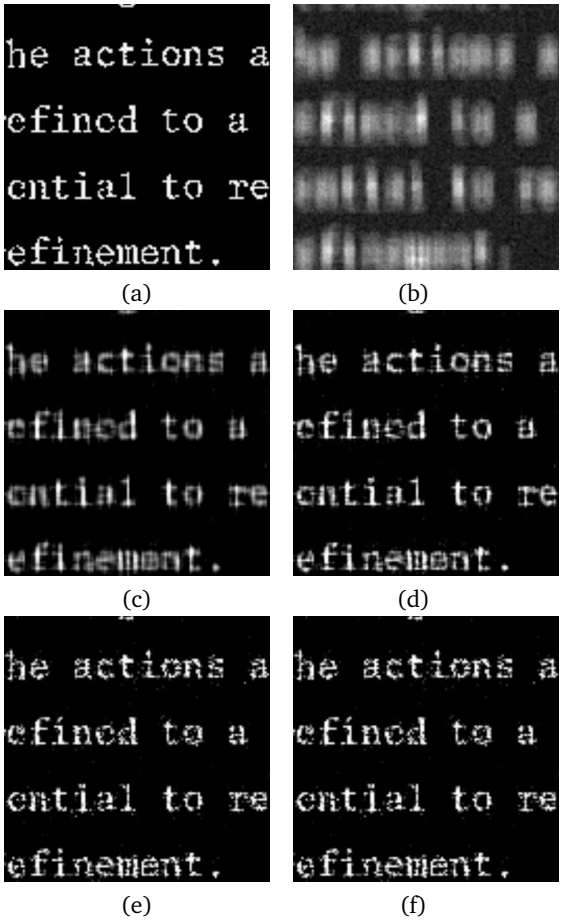


Fig. 2. (a) Original image $\overline{x}$. (b) Degraded image $y$. (c) Image restored by (5) after 50 iterations. (d) Image restored by (6) after 50 iterations. (e) Image restored by (5). (f) Image restored by (6).

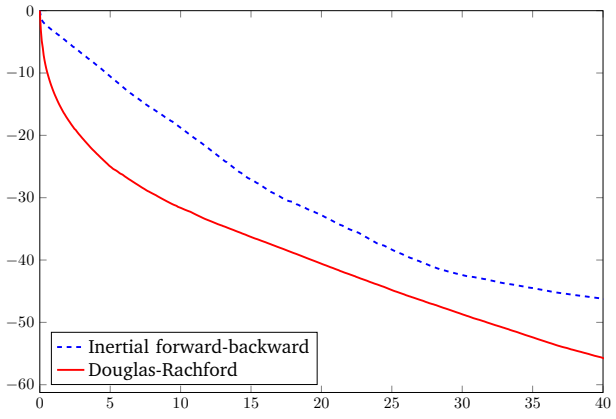


Fig. 3. Normalized distance in dB to the asymptotic image produced by each algorithm versus execution time in seconds.

## IV. Forward-backward-forward splitting

**Problem IV.1** Let $I$ and $J$ be disjoint finite subsets of $\mathbb{N}$ such that $K = I \cup J \neq \varnothing$ and let $f \in \Gamma_0(\mathcal{H})$. For every $k \in K$, suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$ and let $r_k \in \mathcal{G}_k$. For every $i \in I$, let $g_i \in \Gamma_0(\mathcal{G}_i)$ and, for every $j \in J$, let $\mu_j \in ]0, +\infty[$ and let $h_j \colon \mathcal{G}_j \to \mathbb{R}$ be convex and differentiable with a $\mu_j$-Lipschitzian gradient. Assume that

$$0 \in \operatorname{range}\Big(\partial f + \sum_{i\in I} L_i^* \partial g_i L_i + \sum_{j\in J} L_j^* \nabla h_j L_j\Big).$$

The goal is to solve

$$\min_{x\in\mathcal{H}} \; f(x) + \sum_{i\in I} g_i(L_i x) + \sum_{j\in J} h_j(L_j x). \tag{8}$$

Problem IV.1 can be solved by the following algorithm, which is an implementation of the forward-backward-forward algorithm in a primal-dual space [13] (we state only the primal convergence result for brevity).

**Proposition IV.2** [13] *Consider the setting of Problem IV.1. Set* $\beta = \sqrt{\sum_{i\in I}\|L_i\|^2 + \sum_{j\in J}\mu_j\|L_j\|^2}$, *let* $\varepsilon \in ]0, 1/(\beta+1)[$, *let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $[\varepsilon, (1-\varepsilon)/\beta]$, *and let* $(\forall i \in I)\ v_{i,0}^* \in \mathcal{G}_i$. *Let* $x_0 \in \mathcal{H}$ *and iterate*

$$\begin{aligned}
&\textit{for } n = 0, 1, \ldots\\
&\quad y_{1,n} = x_n - \gamma_n\Big(\sum_{i\in I} L_i^* v_{i,n}^* + \sum_{j\in J} L_j^*\big(\nabla h_j(L_j x_n)\big)\Big)\\
&\quad p_n = \operatorname{prox}_{\gamma_n f}\, y_{1,n}\\
&\quad \textit{for every } i \in I\\
&\quad\quad y_{2,i,n} = v_{i,n}^* + \gamma_n L_i x_n\\
&\quad\quad p_{2,i,n} = \operatorname{prox}_{\gamma_n g_i^*}\, y_{2,i,n}\\
&\quad\quad q_{2,i,n} = p_{2,i,n} + \gamma_n L_i p_n\\
&\quad\quad v_{i,n+1}^* = v_{i,n}^* - y_{2,i,n} + q_{2,i,n}\\
&\quad q_{1,n} = p_n - \gamma_n\Big(\sum_{i\in I} L_i^* p_{2,i,n} + \sum_{j\in J} L_j^*\big(\nabla h_j(L_j p_n)\big)\Big)\\
&\quad x_{n+1} = x_n - y_{1,n} + q_{1,n}.
\end{aligned} \tag{9}$$

*Then there exists a solution* $x$ *to* (8) *such that* $x_n \to x$.

We consider the restoration of an $N = 128 \times 128$ image $\overline{x}$ from observations (see Fig. 4(a)–(c)) $y_1 = H_1\overline{x} + w_1$ and $y_2 = H_2\overline{x} + w_2$. Here $H_1$ and $H_2$ model convolution blurs with kernels of size $3\times 11$ and of $7\times 5$, respectively, and $w_1$ and $w_2$ are Gaussian white noise realizations. The blurred image-to-noise-ratios are 27.3 dB and 35.4 dB, respectively. We use $C = [0, 255]^N$ as a hard constraint set. The diffraction of $\overline{x}$ has been observed over some frequency range $R$, possibly with measurement errors. This information is associated with the soft constraint penalty $d_E$, where ($\widehat{x}$ denotes the two-dimensional discrete Fourier transform (DFT)) $E = \big\{x \in \mathbb{R}^N \;\big|\; (\forall k \in R)\ \widehat{x}(k) = \widehat{\overline{x}}(k)\big\}$, with $R$ contains the frequencies in $\{0, \ldots, 15\}^2$ as well as those resulting from the symmetry properties of the DFT. Finally, we use the total variation penalty to control oscillations. This leads to the formulation

$$\min_{x\in C} \; \frac{1}{2}d_E(x) + \frac{2}{5}h(Dx) + \frac{3}{4}\|H_1 x - y_1\|_2^2 + \frac{3}{4}\|H_2 x - y_2\|_2^2, \tag{10}$$

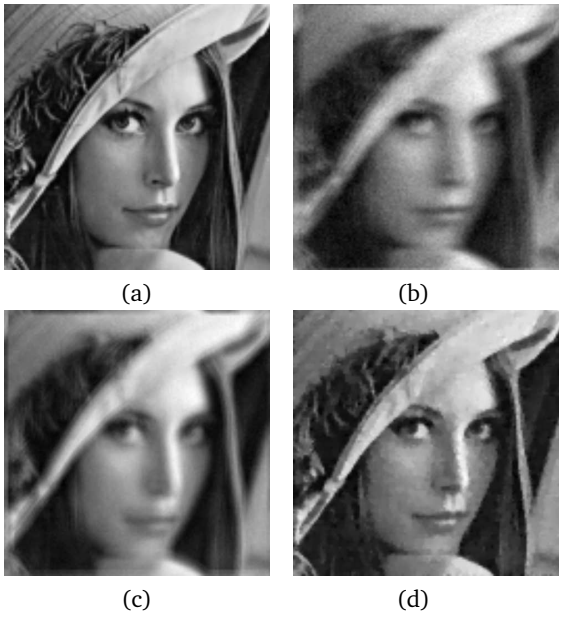

Fig. 4. (a) Original image $\overline{x}$. (b) Degraded image $y_1$. (c) Degraded image $y_2$. (d) Restored image.

where $D\colon \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N \colon x \mapsto (G_1 x, G_2 x)$, $G_1$ and $G_2$ being horizontal and vertical discrete difference operators, and where $(\forall (y_1, y_2) \in \mathbb{R}^N \times \mathbb{R}^N)\ h(y_1, y_2) = \sum_{k=1}^{N} \phi(\|(\eta_{1,k}, \eta_{2,k})\|_2)$, $\phi$ being the standard Huber function (i.e., $\mathcal{H} = \mathbb{R}$ and $C = \{0\}$ in (4)) with parameter $\rho = 2$. We derive from (10) two versions of Problem IV.1.

**Problem IV.3** In Problem IV.1, set $f = \iota_C$, $I = \{1\}$, $g_1 = 0.5 d_E$, $L_1 = \mathrm{Id}$, $J = \{2, 3, 4\}$, $h_2 = 0.4h$, $L_2 = D$, $h_3 = 0.75\| \cdot - y_1\|_2^2$, $L_3 = H_1$, $h_4 = 0.75\| \cdot - y_2\|_2^2$, and $L_4 = H_2$.

**Problem IV.4 (fully proximal)** In Problem IV.1, set $f = \iota_C$, $I = \{1, 2, 3, 4\}$, $J = \varnothing$, $g_1 = 0.5 d_E$, $L_1 = \mathrm{Id}$, $g_2 = 0.4h$, $L_2 = D$, $g_3 = 0.75\|H_1 \cdot - y_1\|_2^2$, $L_3 = \mathrm{Id}$, $g_4 = 0.75\|H_2 \cdot - y_2\|_2^2$, and $L_4 = \mathrm{Id}$.

We consider an implementation of (9) for each problem such that the algorithms are initialized at zero and implemented with the parameter $(\gamma_n)_{n\in\mathbb{N}}$ for which they seem to perform better: $\gamma_n \equiv 0.99/\beta$. The results shown in Figs. 4 and 5 illustrate the better numerical behavior of the fully proximal implementation.

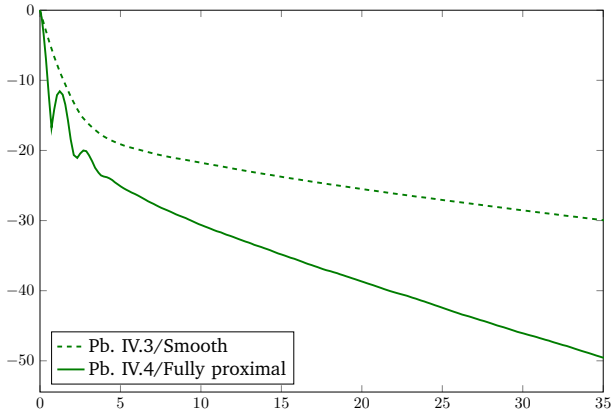


Fig. 5. Normalized distance in dB to the asymptotic image produced by each algorithm versus execution time in seconds.

## References


[1] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot, and L. Laborelli, "Dual block-coordinate forward-backward algorithm with application to deconvolution and deinterlacing of video sequences," *J. Math. Imaging Vision,* vol. 59, pp. 415–431, 2017.

[2] T. Aspelmeier, C. Charitha, and D. R. Luke, "Local linear convergence of the ADMM/Douglas-Rachford algorithms without strong convexity and application to statistical imaging," *SIAM J. Imaging Sci.,* vol. 9, pp. 842–868, 2016.

[3] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed. New York: Springer, 2017.

[4] L. M. Briceño-Arias and P. L. Combettes, "Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery," *Numer. Math. Theory Methods Appl.,* vol. 2, pp. 485–508, 2009.

[5] L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia, and Ph. Ciuciu, "A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging," *Med. Image Anal.,* vol. 15, pp. 185–201, 2011.

[6] A. Chambolle and C. Dossal, "On the convergence of the iterates of the 'Fast iterative shrinkage/thresholding algorithm'," *J. Optim. Theory Appl.,* vol. 166, pp. 968–982, 2015.

[7] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision,* vol. 40, pp. 120–145, 2011.

[8] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems,* vol. 23, pp. 1495–1518, 2007.

[9] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A convex approach for image restoration with exact Poisson-Gaussian likelihood," *SIAM J. Imaging Sci.,* vol. 8, pp. 2662–2682, 2015.

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

[11] P. L. Combettes and L. E. Glaudin, "Proximal activation of smooth functions in splitting algorithms for convex image recovery," arXiv:1803.02919

[12] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering,* pp. 185–212. Springer, New York, 2011.

[13] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal.*, vol. 20, pp. 307–330, 2012.

[14] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.,* vol. 4, pp. 1168–1200, 2005.

[15] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris,* vol. A255, pp. 2897–2899, 1962.

[16] D. O'Connor and L. Vandenberghe, "Primal-dual decomposition by operator splitting and applications to image deblurring," *SIAM J. Imaging Sci.,* vol. 7, pp. 1724–1754, 2014.

[17] N. Papadakis, G. Peyré, and E. Oudet, "Optimal transport with proximal splitting," *SIAM J. Imaging Sci.,* vol. 7, pp. 212–238, 2014.

[18] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Trans. Circuits Syst.,* vol. 25, pp. 694–702, 1978.

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