Proximal Activation of Smooth Functions in Splitting Algorithms for Convex Image Recovery*

Patrick L. Combettes† and Lilian E. Glaudin‡

Abstract. Structured convex optimization problems 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. A novel variational model to relax inconsistent convex feasibility problems is also investigated within the proposed framework. Several numerical applications to image recovery are presented to compare the behavior of fully proximal versus mixed proximal/gradient implementations of several splitting algorithms.

Key words. convex optimization, image recovery, inconsistent convex feasibility problem, proximal splitting algorithm, proximity operator

1. Introduction.

Splitting in convex optimization methods for image recovery can be traced back to the influential work of Youla [62, 64]. The convex feasibility framework he proposed consists in formulating the image recovery problem as that of finding an image in a Hilbert 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

$$\text{find } x \in \bigcap_{i=1}^{m} C_i. \tag{1.1}$$

Now, for every $i \in \{1,\ldots,m\}$, let $\operatorname{proj}_{C_i}$ be the projection operator onto C_i, which maps each $x \in \mathcal{H}$ to its unique closest point in C_i, that is,

$$\operatorname{proj}_{C_i}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname*{argmin}_{y\in\mathcal{H}} \bigg(\iota_{C_i}(y) + \frac{1}{2}\|x-y\|^2\bigg), \tag{1.2}$$

$$\text{where} \quad \iota_{C_i}\colon y \mapsto \begin{cases} 0, & \text{if } y \in C_i;\\ +\infty, & \text{if } y \notin C_i. \end{cases}$$

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 [8, 23]. For instance, the POCS (Projection Onto Convex Sets) algorithm advocated in [64] is governed

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

†North Carolina State University, Department of Mathematics, Raleigh, NC 27695-8205, USA (plc@math.ncsu.edu)

‡Sorbonne Université, Laboratoire Jacques-Louis Lions, F-75005 Paris, France (glaudin@ljll.math.upmc.fr)

by the updating rule

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (\operatorname{proj}_{C_1} \circ \cdots \circ \operatorname{proj}_{C_m}) x_n. \tag{1.3}$$

Convex variational formulations arising in modern image recovery have complex structures that require sophisticated analysis tools and solution methods. Since projection operators are of limited use beyond feasibility and best approximation problems, to solve such formulations, one strategy is to use an extended notion of a projection operator. In [35] it was suggested to use Moreau's proximity operator [51] for this purpose. Recall that the proximity operator of a proper lower semicontinuous convex function $\varphi\colon \mathcal{H} \to]-\infty, +\infty]$ is

$$\operatorname{prox}_{\varphi}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname*{argmin}_{y \in \mathcal{H}} \left(\varphi(y) + \frac{1}{2}\|x - y\|^2 \right), \tag{1.4}$$

and that it reduces to (1.2) when $\varphi = \iota_{C_i}$. We refer the reader to [9, Chapter 24] for a detailed account of the properties of proximity operators with various examples, to [31] for a tutorial on proximal methods in signal processing, and to [5, 11, 19, 21, 30, 52, 53, 54] for specific applications to image recovery. Current proximal splitting methods can handle highly structured convex minimization problems such as the following, which will be the focus of our discussion (see below for notation).

Problem 1.1. Let $\mathcal{H}$ be a real Hilbert space, let I and J be disjoint finite subsets of $\mathbb{N}$ such that $K = I \cup J \neq \varnothing$, let $f \in \Gamma_0(\mathcal{H})$, and let $(\mathcal{G}_k)_{k \in K}$ be a family of real Hilbert spaces. For every $k \in K$, suppose that $L_k\colon \mathcal{H} \to \mathcal{G}_k$ is a nonzero bounded linear operator. 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 μ_j-Lipschitzian gradient. Assume that (see [32, Proposition 4.3] for sufficient conditions)

$$0 \in \operatorname{range}\left(\partial f + \sum_{i \in I} L_i^* \circ \partial g_i \circ L_i + \sum_{j \in J} L_j^* \circ (\nabla h_j) \circ L_j \right). \tag{1.5}$$

The goal is to

$$\operatorname*{minimize}_{x \in \mathcal{H}} \; f(x) + \sum_{i \in I} g_i(L_i x) + \sum_{j \in J} h_j(L_j x). \tag{1.6}$$

The principle of a splitting method for solving (1.6) is to use separately each of the functions f, $(g_i)_{i \in I}$, and $(h_j)_{j \in J}$, and each of the operators $(L_k)_{k \in K}$, so as to reduce the execution of the algorithm to a sequence of simple steps. A prevalent viewpoint in first order convex splitting algorithms is that to activate each function φ appearing in the model there are two options:

- if φ is smooth, i.e., real-valued and differentiable everywhere with a Lipschitzian gradient, then use $\nabla\varphi$;
- otherwise, use φ proximally, i.e., via its proximity operator (1.4).

In the present paper we propose a more nuanced viewpoint and submit that, when φ is smooth, it may be computationally advantageous to activate it proximally when its proximity

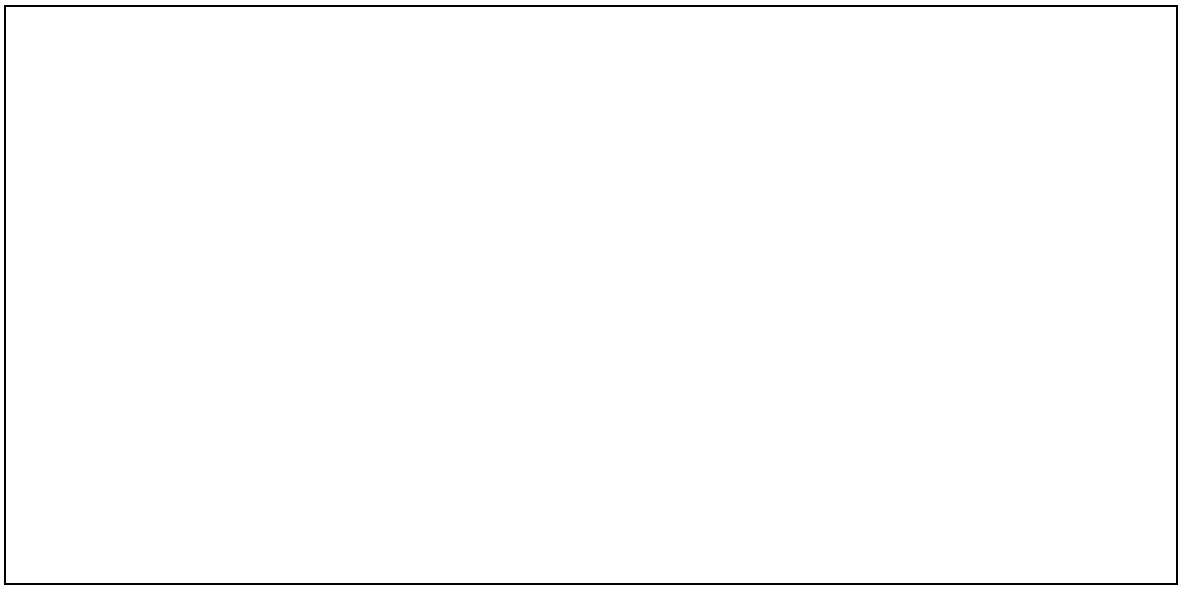

Figure 1.1. *Comparison of the steepest descent method (1.8) (in green) and of the proximal point algorithm (1.9) (in red) in $\mathcal{H} = \mathbb{R}^2$ for $\varphi\colon (\xi_1, \xi_2) \mapsto 9\xi_1^2 - 14\xi_1\xi_2 + 9\xi_2^2$. The ellipsoids represent the level lines of φ. The steepest descent 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 steepest descent, 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 steepest descent method obtained by setting $f = 0$ and $h = \varphi$ in Algorithm 3.3, and choosing the parameters $\alpha = 2.01$ and $\gamma = 1/\beta$, which gave the fastest convergence.*

operator can be implemented. To motivate this viewpoint, let us first observe that a tight Lipschitz constant for the gradient of φ may not be easy to estimate (see, e.g., [1, 13, 17]), 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 performance of algorithms. To illustrate this fact, consider the problem of minimizing a differentiable convex function $\varphi\colon \mathcal{H} \to \mathbb{R}$ with a β-Lipschitzian gradient (see Fig. 1.1 for a concrete example). The associated continuous-time gradient dynamics is [6, Section 3.4]

$$x(0) = x_0 \quad \text{and} \quad -\frac{dx(t)}{dt} = \nabla\varphi(x(t)). \tag{1.7}$$

The forward Euler (explicit) discretization of this equation with time step $\gamma \in \left]0, +\infty\right[$ assumes the form $-(x_{n+1} - x_n)/\gamma = \nabla\varphi(x_n)$, which leads to the steepest descent algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma\nabla\varphi(x_n). \tag{1.8}$$

On the other hand, the backward Euler (implicit) discretization of (1.7) is $-(x_{n+1} - x_n)/\gamma =$

$\nabla\varphi(x_{n+1})$, which leads to Martinet's proximal point algorithm [49]

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma\varphi} x_n. \tag{1.9}$$

Alternatively, it follows from [9, Proposition 12.30] that the proximal point algorithm coincides with the steepest descent method applied to the Moreau envelope of φ, namely,

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma\varphi} x_n = x_n - \gamma\nabla\big({}^{\gamma}\varphi\big)(x_n), \quad \text{where} \quad {}^{\gamma}\varphi = \varphi \,\square\, (q/\gamma). \tag{1.10}$$

While the convergence of (1.9) is guaranteed for any $\gamma \in \left]0, +\infty\right[$, that of (1.8) requires $\gamma < 2/\beta$ [9, Chapter 28], which results in potentially slow convergence. Historically, the idea of using proximal steps in smooth minimization problems can be found in [12, Section 5.8]. There, the problem under consideration is the standard least-squares problem of minimizing the smooth function $\varphi\colon \mathbb{R}^N \to \mathbb{R}\colon x \mapsto \|Ax-b\|^2/2$ in connection with the numerical inversion of the Laplace transform. Given $\gamma \in \left]0, +\infty\right[$ and $x_0 \in \mathbb{R}^N$, the algorithm proposed in [12, Eq. (5.8.3)] is

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \big(\operatorname{Id} + \gamma A^\top A\big)^{-1}\big(x_n + \gamma A^\top b\big), \tag{1.11}$$

and it is reported to be better than the standard steepest descent approach. Remarkably, (1.11) is nothing but an early instance of the proximal point algorithm (1.9).

The paper is organized as follows. In Section 2, we enrich the list of known proximity operators by providing new closed form expressions for those of various smooth convex functions commonly encountered in applications. This investigation is of interest in its own right since some splitting algorithms operate exclusively with proximal steps; see, e.g., [19, 27, 28, 30]. In connection with the numerical solution of Problem 1.1, we review in Section 3 some pertinent proximal splitting methods. Image recovery applications are presented in Section 4. Numerical comparisons between splitting algorithms in which smooth functions are activated via gradient steps and those in which all functions are activated via their proximity operators are conducted. In particular, in Section 4.4, we propose a new variational model, based on Problem 1.1 and the results of Section 2, to relax inconsistent convex feasibility problems. 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 [9]. Throughout, $\mathcal{H}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$, associated norm $\|\cdot\|$, and identity operator Id. Weak convergence is denoted by $\rightharpoonup$. Given a real Hilbert space $\mathcal{G}$, we denote by $\mathcal{B}(\mathcal{H}, \mathcal{G})$ the space of continuous linear operators from $\mathcal{H}$ to $\mathcal{G}$. We set $q = \|\cdot\|^2/2$ and denote by $\Gamma_0(\mathcal{H})$ the class of lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ such that $\operatorname{dom} f = \big\{x \in \mathcal{H} \;\big|\; f(x) < +\infty\big\} \neq \varnothing$. Let $f \in \Gamma_0(\mathcal{H})$. Then f^* denotes the conjugate of f, ∂f the subdifferential of f, and $f \,\square\, g$ the inf-convolution of f and $g \in \Gamma_0(\mathcal{H})$. Let C be a convex subset of $\mathcal{H}$. The strong relative interior of C is denoted by $\operatorname{sri} C$, the indicator function of C by ι_C, the distance function to C by d_C, the support function of C by σ_C, and, if C is nonempty and closed, the projection operator onto C by proj_C. The Hilbert direct sum of family of real Hilbert spaces $(\mathcal{H}_i)_{i\in I}$ is denoted by $\bigoplus_{i\in I} \mathcal{H}_i$; in addition if, for every $i \in I$, $f_i\colon \mathcal{H}_i \to [0, +\infty]$, then

$$\bigoplus_{i\in I} f_i\colon \bigoplus_{i\in I} \mathcal{H}_i \to [0, +\infty]\colon (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i). \tag{1.12}$$

The standard Euclidean norm on $\mathbb{R}^N$ is denoted by $\|\cdot\|_2$.

2. Proximity operators of smooth convex functions. Let $\beta \in]0,+\infty[$, let $\gamma \in]0,+\infty[$, and let $h\colon \mathcal{H} \to \mathbb{R}$ be a convex function with a β-Lipschitzian gradient. Then there exists a function $g \in \Gamma_0(\mathcal{H})$ such that $h = g^* \,\square\, (\beta q)$ [9, Corollary 18.19]. In this case, we derive from [9, Propositions 12.30 and 24.8(vii)] that

$$\nabla h = \beta\big(\mathrm{Id} - \mathrm{prox}_{g^*/\beta}\big) \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = \mathrm{Id} + \frac{\gamma\beta}{\gamma\beta+1}\big(\mathrm{prox}_{(\gamma\beta+1)g^*/\beta} - \mathrm{Id}\big). \tag{2.1}$$

The closed form expression for $\mathrm{prox}_{\gamma h} x$ above is, however, of limited use since g^* and its proximity operator are usually not available explicitly. Even when $\mathcal{H} = \mathbb{R}$, computing the proximity operator of a smooth convex function may be involved: for instance the derivative of $h\colon x \mapsto \sqrt[3]{1+x^6}/3$ is $\sqrt[3]{2}$-Lipschitzian but evaluating $\mathrm{prox}_{\gamma h}$ requires solving a high degree polynomial equation. Nonetheless, as we now show, a variety of smooth convex functions encountered in applications have readily computable proximity operators.

2.1. Functions involving distances. The following fact will be needed.

Lemma 2.1. [13, Proposition 2.1] *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi \in \Gamma_0(\mathbb{R})$ be even, and set $\varphi = \phi \circ d_C$. Then $\varphi \in \Gamma_0(\mathcal{H})$. Moreover, $\mathrm{prox}_\varphi = \mathrm{proj}_C$ if $\mathrm{dom}\,\phi = \{0\}$ and, otherwise, for every $x \in \mathcal{H}$,*

$$\mathrm{prox}_\varphi x = \begin{cases} x + \dfrac{\mathrm{prox}_{\phi^*} d_C(x)}{d_C(x)}\big(\mathrm{proj}_C x - x\big), & \text{if } \ d_C(x) > \max \partial\phi(0); \\ \mathrm{proj}_C x, & \text{if } \ x \notin C \text{ and } \ d_C(x) \leqslant \max \partial\phi(0); \\ x, & \text{if } \ x \in C. \end{cases} \tag{2.2}$$

We start with an example which leads to affine gradient and proximal operators.

Example 2.2. 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 $h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\sum_{i\in I} \alpha_i d^2_{V_i}(L_i x - r_i)$ and $Q = (\mathrm{Id} + \gamma \sum_{i\in I} \alpha_i L_i^* \mathrm{proj}_{V_i^\perp} L_i)^{-1}$. 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 Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \sum_{i\in I} \alpha_i L_i^*\Big(\mathrm{proj}_{V_i^\perp}\big(L_i x - r_i\big)\Big), \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = Q\bigg(x + \gamma \sum_{i\in I} \alpha_i L_i^*\Big(\mathrm{proj}_{V_i^\perp} r_i\Big)\bigg). \tag{2.3}$$

Proof. The convexity of h is clear. We have $h(x) = (1/2)\sum_{i\in I} \alpha_i \|\mathrm{proj}_{V_i^\perp}(L_i x - r_i)\|^2$ and $\nabla h(x)$ is given by (2.3) since $(\forall i \in I)\ \nabla d^2_{V_i}/2 = \mathrm{Id} - \mathrm{proj}_{V_i} = \mathrm{proj}_{V_i^\perp}$. Moreover,

$$(\forall i \in I) \quad \|L_i^* \mathrm{proj}_{V_i^\perp} L_i\| \leqslant \|L_i^*\|\,\|\mathrm{proj}_{V_i^\perp}\|\,\|L_i\| \leqslant \|L_i\|^2. \tag{2.4}$$

Hence, for every $y \in \mathcal{H}$,

$$\|\nabla h(x) - \nabla h(y)\| = \bigg\| \sum_{i\in I} \alpha_i L_i^*\Big(\mathrm{proj}_{V_i^\perp} L_i(x-y)\Big)\bigg\| \leqslant \sum_{i\in I} \alpha_i \|L_i\|^2 \|x-y\| = \beta\|x-y\|. \tag{2.5}$$

Now set $p = \operatorname{prox}_{\gamma h} x$. Then we derive from (2.3) that

$$x - p = \gamma \nabla h(p) = \gamma \left(\sum_{i \in I} \alpha_i L_i^* \operatorname{proj}_{V_i^\perp} L_i \right) p - \gamma \sum_{i \in I} \alpha_i L_i^* \left(\operatorname{proj}_{V_i^\perp} r_i \right), \tag{2.6}$$

which yields the expression for $\operatorname{prox}_{\gamma h} x$. ∎

The next construction, which involves the distance function d_C to a convex set C, will be seen to capture a broad range of functions of interest.

Example 2.3. 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 β-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 Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\phi'\big(d_C(x)\big)}{d_C(x)} \big(x - \operatorname{proj}_C x\big), & \text{if } \ x \notin C; \\ 0, & \text{if } \ x \in C, \end{cases} \tag{2.7}$$

and

$$\operatorname{prox}_{\gamma h} x = \begin{cases} \operatorname{proj}_C x + \dfrac{\operatorname{prox}_{\gamma \phi} d_C(x)}{d_C(x)} (x - \operatorname{proj}_C x), & \text{if } \ x \notin C; \\ x, & \text{if } \ x \in C. \end{cases} \tag{2.8}$$

Proof. Since ϕ and d_C are convex and ϕ is increasing on $[0, +\infty[$, h is convex [9, Proposition 11.7(ii)]. In addition, since [9, Proposition 11.7(i)] asserts that 0 is a minimizer of ϕ,

$$\partial \phi(0) = \{\phi'(0)\} = \{0\}. \tag{2.9}$$

First, we derive (2.7) from [9, Proposition 17.33(ii)] and (2.9). Next, we infer from [9, Proposition 13.26] that $h^* = \sigma_C + \phi^* \circ \| \cdot \|$. We invoke [9, Theorem 18.15] to deduce that ϕ^* is $(1/\beta)$-strongly convex and that h^* is therefore likewise, and then to conclude that ∇h is β-Lipschitzian. On the other hand, we derive from Lemma 2.1, (2.9), and Moreau's decomposition [9, Remark 14.4] that

$$\begin{aligned} \operatorname{prox}_{\gamma h} x &= \begin{cases} x + \dfrac{\operatorname{prox}_{\phi^*} d_C(x)}{d_C(x)} \big(\operatorname{proj}_C x - x\big), & \text{if } \ d_C(x) > \max \partial \phi(0); \\ \operatorname{proj}_C x, & \text{if } \ d_C(x) \leqslant \max \partial \phi(0) \end{cases} \\ &= \begin{cases} x + \dfrac{d_C(x) - \operatorname{prox}_{\gamma \phi} d_C(x)}{d_C(x)} (\operatorname{proj}_C x - x), & \text{if } \ d_C(x) > 0; \\ \operatorname{proj}_C x, & \text{if } \ d_C(x) \leqslant 0 \end{cases} \end{aligned} \tag{2.10}$$

$$= \begin{cases} \operatorname{proj}_C x + \dfrac{\operatorname{prox}_{\gamma \phi} d_C(x)}{d_C(x)} (x - \operatorname{proj}_C x), & \text{if } \ x \notin C; \\ x, & \text{if } \ x \in C, \end{cases} \tag{2.11}$$

as announced. ∎

We now investigate a generalization of the Vapnik ε-insensitive loss function [59].

Example 2.4 (abstract smooth Vapnik loss function). Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\varepsilon \in]0,+\infty[$, let $\beta \in]0,+\infty[$, let $\psi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a β-Lipschitzian derivative, and set $h = \psi \circ \max(d_C - \varepsilon, 0)$. Let $\gamma \in]0,+\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\psi'\big(d_C(x) - \varepsilon\big)}{d_C(x)}\big(x - \operatorname{proj}_C x\big), & \text{if } \ d_C(x) > \varepsilon; \\ 0, & \text{if } \ d_C(x) \leqslant \varepsilon, \end{cases} \tag{2.12}$$

and

$$\operatorname{prox}_{\gamma h} x = \begin{cases} \operatorname{proj}_C x + \dfrac{\varepsilon + \operatorname{prox}_{\gamma\psi}\big(d_C(x) - \varepsilon\big)}{d_C(x)}(x - \operatorname{proj}_C x), & \text{if } \ d_C(x) > \varepsilon; \\ x, & \text{if } \ d_C(x) \leqslant \varepsilon. \end{cases} \tag{2.13}$$

Proof. Let $\vartheta = \max(|\cdot| - \varepsilon, 0)$ be the standard Vapnik loss function, set $\phi = \psi \circ \vartheta$, and let $\xi \in \mathbb{R}$. Upon applying Example 2.3 in $\mathbb{R}$ with $C = [-\varepsilon, \varepsilon]$, we obtain that $\phi\colon \mathbb{R} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian derivative, that

$$\phi'(\xi) = \begin{cases} \psi'(|\xi| - \varepsilon)\operatorname{sign}(\xi), & \text{if } \ |\xi| > \varepsilon; \\ 0, & \text{if } \ |\xi| \leqslant \varepsilon, \end{cases} \tag{2.14}$$

and that

$$\operatorname{prox}_{\gamma\phi}\xi = \begin{cases} \big(\varepsilon + \operatorname{prox}_{\gamma\psi}(|\xi| - \varepsilon)\big)\operatorname{sign}(\xi), & \text{if } \ |\xi| > \varepsilon; \\ \xi, & \text{if } \ |\xi| \leqslant \varepsilon. \end{cases} \tag{2.15}$$

Since $h = \phi \circ d_C$ and ϕ is even, we apply Example 2.3 to conclude. ∎

The following is an extension of the Huber loss function [42].

Example 2.5 (abstract Huber function). Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\rho \in]0,+\infty[$, and set

$$h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \begin{cases} \rho d_C(x) - \dfrac{\rho^2}{2}, & \text{if } \ d_C(x) > \rho; \\ \dfrac{d_C(x)^2}{2}, & \text{if } \ d_C(x) \leqslant \rho. \end{cases} \tag{2.16}$$

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

$$\nabla h(x) = \begin{cases} \dfrac{\rho}{d_C(x)}\big(x - \operatorname{proj}_C x\big), & \text{if } \ d_C(x) > \rho; \\ x - \operatorname{proj}_C x, & \text{if } \ d_C(x) \leqslant \rho, \end{cases} \tag{2.17}$$

and

$$\operatorname{prox}_{\gamma h} x = \begin{cases} x + \dfrac{\gamma\rho}{d_C(x)}\big(\operatorname{proj}_C x - x\big), & \text{if } \ d_C(x) > (\gamma+1)\rho; \\ \dfrac{1}{\gamma+1}\big(x + \gamma \operatorname{proj}_C x\big), & \text{if } \ d_C(x) \leqslant (\gamma+1)\rho. \end{cases} \tag{2.18}$$

Proof. Let

$$\mathfrak{h}_\rho \colon \mathbb{R} \to \mathbb{R} \colon \xi \mapsto \begin{cases} \rho|\xi| - \dfrac{\rho^2}{2}, & \text{if } \ |\xi| > \rho; \\ \dfrac{|\xi|^2}{2}, & \text{if } \ |\xi| \leqslant \rho \end{cases} \tag{2.19}$$

be the standard Huber function with parameter ρ. Then $\mathfrak{h}'_\rho$ is 1-Lipschitzian. In addition, using the expression of $\operatorname{prox}_{\gamma\mathfrak{h}_\rho}$ from [9, Example 24.9] and then Example 2.3, we obtain (2.18). ■

Example 2.6. Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\omega \in \left]0, +\infty\right[$, and set $\beta = \omega^2$ and $h = \omega d_C - \ln(1 + \omega d_C)$. Let $\gamma \in \left]0, +\infty\right[$ and let $x \in \mathcal{H}$. Then $h \colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\omega^2}{1 + \omega d_C(x)}\big(x - \operatorname{proj}_C x\big), & \text{if } \ x \notin C; \\ 0, & \text{if } \ x \in C, \end{cases} \tag{2.20}$$

and

$$\operatorname{prox}_{\gamma h} x = \tag{2.21}$$
$$\begin{cases} \operatorname{proj}_C x + \dfrac{\gamma\omega^2 + 1 - \omega d_C(x) - \sqrt{|\omega d_C(x) - \gamma\omega^2 - 1|^2 + 4\omega d_C(x)}}{2\omega d_C(x)}(\operatorname{proj}_C x - x), & \text{if } \ x \notin C; \\ x, & \text{if } \ x \in C. \end{cases}$$

Proof. We apply Example 2.3 with $\phi = \omega|\cdot| - \ln(1 + \omega|\cdot|)$. Note that $\phi' \colon \xi \mapsto \omega^2\xi/(1 + \omega|\xi|)$ is ω^2-Lipschitzian. Furthermore, we derive $\operatorname{prox}_{\gamma\phi}$ by arguing as in [9, Example 24.42] (where $\gamma = 1$) and we then invoke (2.8) to get (2.21). ■

The following extension of Example 2.3 involves a composition with a linear operator.

Example 2.7. Let $\mathcal{G}$ be a real Hilbert space and let $M \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that $MM^* = \theta \operatorname{Id}$ for some $\theta \in \left]0, +\infty\right[$. Let D be a nonempty closed convex subset of $\mathcal{G}$, let $\mu \in \left]0, +\infty\right[$, let $\phi \colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a μ-Lipschitzian derivative, and set $h = \phi \circ d_D \circ M$ and $\beta = \mu\|M\|^2$. Let $\gamma \in \left]0, +\infty\right[$ and $x \in \mathcal{H}$. Then $h \colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\phi'\big(d_D(Mx)\big)}{d_D(Mx)} M^*\big(Mx - \operatorname{proj}_D(Mx)\big), & \text{if } \ Mx \notin D; \\ 0, & \text{if } \ Mx \in D, \end{cases} \tag{2.22}$$

and

(2.23)
$$\operatorname{prox}_{\gamma h} x = \begin{cases} x + \dfrac{\theta^{-1}\big(d_D(Mx) - \operatorname{prox}_{\gamma\theta\phi} d_D(Mx)\big)}{d_D(Mx)} M^*\big(\operatorname{proj}_D(Mx) - Mx\big), & \text{if } Mx \notin D; \\ x, & \text{if } Mx \in D. \end{cases}$$

Proof. We have $h = (\phi \circ d_D) \circ M$ and therefore $\nabla h = M^* \circ \nabla(\phi \circ d_D) \circ M$. In turn, the Lipschitz constant of ∇h is $\|M^*\|\mu\|M\| = \beta$, and we derive (2.22) from (2.7). Now set $g = \gamma\phi \circ d_D$. We derive from [9, Proposition 24.14] that

$$\operatorname{prox}_{\gamma h} x = \operatorname{prox}_{g \circ M} x = x + \theta^{-1} M^*\big(\operatorname{prox}_{\theta g}(Mx) - Mx\big). \tag{2.24}$$

We then obtain the expression for $\operatorname{prox}_{\theta g} = \operatorname{prox}_{\gamma\theta\phi\circ d_D}$ from (2.8), which yields (2.23). ∎

Remark 2.8. The condition $MM^* = \theta\,\mathrm{Id}$ used in Example 2.7 arises in particular in problems involving tight frame representations [20]. 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; see Propositions 3.8, 3.10, and 3.12 below and [9]. One can then invoke Example 2.3 to compute the former.

2.2. Integral functions.

Example 2.9. Let $(\Omega, \mathcal{F}, \mu)$ be a complete σ-finite measure space, let $(\mathsf{H}, \langle\cdot \mid \cdot\rangle_{\mathsf{H}})$ be a separable real Hilbert space, let C be a closed convex subset of H such that $0 \in \mathsf{C}$, and let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a β-Lipschitzian derivative. Suppose that $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{H})$, and that $\mu(\Omega) < +\infty$ or $\phi(0) = 0$. Set $h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \int_\Omega \phi(d_{\mathsf{C}}(x(\omega)))\mu(d\omega)$. Let $\gamma \in \left]0, +\infty\right[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient, and for μ-almost every $w \in \Omega$,

$$\big(\nabla h(x)\big)(\omega) = \begin{cases} \dfrac{\phi'(d_{\mathsf{C}}(x(\omega)))}{d_{\mathsf{C}}(x(\omega))}\big(x(\omega) - \operatorname{proj}_{\mathsf{C}} x(\omega)\big), & \text{if } x(\omega) \notin \mathsf{C}; \\ 0, & \text{if } x(\omega) \in \mathsf{C}, \end{cases} \tag{2.25}$$

and

$$\big(\operatorname{prox}_{\gamma h} x\big)(\omega) = \begin{cases} \operatorname{proj}_{\mathsf{C}} x(\omega) + \dfrac{\operatorname{prox}_{\gamma\phi} d_{\mathsf{C}}(x(\omega))}{d_{\mathsf{C}}(x(\omega))}\big(x(\omega) - \operatorname{proj}_{\mathsf{C}} x(\omega)\big), & \text{if } x(\omega) \notin \mathsf{C}; \\ x(\omega), & \text{if } x(\omega) \in \mathsf{C}. \end{cases} \tag{2.26}$$

Proof. Set $\varphi = \phi \circ d_{\mathsf{C}}$. As seen in Example 2.3, $\varphi\colon \mathsf{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \nabla\varphi(\mathsf{x}) = \begin{cases} \dfrac{\phi'\big(d_{\mathsf{C}}(\mathsf{x})\big)}{d_{\mathsf{C}}(\mathsf{x})}\big(\mathsf{x} - \operatorname{proj}_{\mathsf{C}}\mathsf{x}\big), & \text{if } \mathsf{x} \notin \mathsf{C}; \\ 0, & \text{if } \mathsf{x} \in \mathsf{C}, \end{cases} \tag{2.27}$$

and

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \operatorname{prox}_{\gamma\varphi}\mathsf{x} = \begin{cases} \operatorname{proj}_{\mathsf{C}}\mathsf{x} + \dfrac{\operatorname{prox}_{\gamma\phi} d_{\mathsf{C}}(\mathsf{x})}{d_{\mathsf{C}}(\mathsf{x})}(\mathsf{x} - \operatorname{proj}_{\mathsf{C}}\mathsf{x}), & \text{if } \mathsf{x} \notin \mathsf{C}; \\ \mathsf{x}, & \text{if } \mathsf{x} \in \mathsf{C}. \end{cases} \tag{2.28}$$

Since ϕ is convex and even, it is minimized by 0 [9, Proposition 11.7(i)]. Hence, $\phi'(0) = 0$ and, since $0 \in \mathsf{C}$, $\varphi(0) = \phi(0)$, while (2.27) yields $\nabla\varphi(0) = 0$. Consequently, by virtue of the descent lemma [9, Theorem 18.15(iii)],

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \varphi(\mathsf{x}) \leqslant \varphi(0) + \langle \mathsf{x} \mid \nabla\varphi(0)\rangle_{\mathsf{H}} + \frac{\beta}{2}\|\mathsf{x}\|_{\mathsf{H}}^2 = \varphi(0) + \frac{\beta}{2}\|\mathsf{x}\|_{\mathsf{H}}^2. \tag{2.29}$$

In turn,

$$h(x) = \int_\Omega \varphi(x(\omega))\mu(d\omega) \leqslant \varphi(0)\mu(\Omega) + \frac{\beta}{2}\|x\|^2 < +\infty. \tag{2.30}$$

On the other hand, [9, Proposition 16.63(ii)] asserts that $\nabla h(x) = (\nabla\varphi) \circ x$ μ-a.e. which, combined with (2.27), yields (2.25). Now let y be in $\mathcal{H}$. Then

$$\begin{aligned} \|\nabla h(x) - \nabla h(y)\|^2 &= \int_\Omega \|(\nabla h(x))(\omega) - (\nabla h(y))(\omega)\|_{\mathsf{H}}^2 \mu(d\omega) \\ &= \int_\Omega \|\nabla\varphi(x(\omega)) - \nabla\varphi(y(\omega))\|_{\mathsf{H}}^2 \mu(d\omega) \\ &\leqslant \beta^2 \int_\Omega \|x(\omega) - y(\omega)\|_{\mathsf{H}}^2 \mu(d\omega) \\ &= \beta^2\|x - y\|^2, \end{aligned} \tag{2.31}$$

which shows that ∇h is β-Lipschitzian. Finally, we apply [9, Proposition 24.13] to derive (2.26) from (2.28). ∎

Remark 2.10. Let Ω be a nonempty bounded smooth open subset of $\mathbb{R}^2$, let $\mathsf{H} = \mathbb{R}^2$, let μ be the Lebesgue measure, and suppose that $\mathsf{C} = \{0\}$ in Example 2.9. Furthermore, let $\mathfrak{h}_\rho$ be the Huber function of (2.19) and, for every $x \in H_0^1(\Omega)$, let Dx be the gradient of x. Then the function

$$h \circ D\colon H_0^1(\Omega) \to \mathbb{R}\colon x \mapsto \int_\Omega \mathfrak{h}_\rho(\|Dx(\omega)\|_2)d\omega \tag{2.32}$$

can be found in [41, 44, 48, 55] and it is called the Gauss-TV (or TV-Huber) function.

2.3. Functionals involving orthonormal decompositions. We first revisit a construction proposed in [35]; see also [33, 37, 38] for special cases.

Example 2.11. 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 $\beta_k \in \left]0, +\infty\right[$ 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 ϕ_k' is β_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 Fréchet differentiable with a β-Lipschitzian gradient,

$$(\forall x \in \mathcal{H}) \quad \nabla h(x) = \sum_{k\in\mathbb{K}} \phi_k'(\langle x \mid e_k\rangle)e_k, \tag{2.33}$$

and

$$(\forall x \in \mathcal{H}) \quad \operatorname{prox}_{\gamma h} x = \sum_{k\in\mathbb{K}} \big(\operatorname{prox}_{\gamma\phi_k}\langle x \mid e_k\rangle\big)e_k. \tag{2.34}$$

Proof. The identity (2.34) is established in [35]. The frame analysis operator is

$$F\colon \mathcal{H} \to \ell^2(\mathbb{K})\colon x \mapsto (\langle x \mid e_k\rangle)_{k\in\mathbb{K}} \tag{2.35}$$

and its adjoint is the frame synthesis operator

$$F^*\colon \ell^2(\mathbb{K}) \to \mathcal{H}\colon (\xi_k)_{k\in\mathbb{K}} \mapsto \sum_{k\in\mathbb{K}} \xi_k e_k. \tag{2.36}$$

Now denote by $\mathsf{x} = (\xi_k)_{k\in\mathbb{K}}$ a generic element in $\ell^2(\mathbb{K})$ and define

$$\varphi\colon \ell^2(\mathbb{K}) \to]-\infty, +\infty] : \mathsf{x} \mapsto \sum_{k\in\mathbb{K}} \phi_k(\xi_k). \tag{2.37}$$

Then $h = \varphi \circ F$. Since all the functions $(\phi_k)_{k\in\mathbb{K}}$ are minimized at 0, we have $(\forall k \in \mathbb{K})$ $\phi_k'(0) = 0 = \phi_k(0)$. In turn, we derive from the descent lemma [9, Theorem 18.15(iii)] that

$$(\forall k \in \mathbb{K})(\forall \xi_k \in \mathbb{R}) \quad \phi_k(\xi_k) \leqslant \phi_k(0) + (\xi_k - 0)\phi_k'(0) + \frac{\beta_k}{2}|0 - \xi_k|^2 = \frac{\beta}{2}|\xi_k|^2. \tag{2.38}$$

As a result,

$$\big(\forall \mathsf{x} \in \ell^2(\mathbb{K})\big) \quad \varphi(\mathsf{x}) = \sum_{k\in\mathbb{K}} \phi_k(\xi_k) \leqslant \frac{\beta}{2}\sum_{k\in\mathbb{K}} |\xi_k|^2 = \frac{\beta}{2}\|\mathsf{x}\|^2 \tag{2.39}$$

and, therefore, $\varphi\colon \ell^2(\mathbb{K}) \to \mathbb{R}$. In addition,

$$(\forall k \in \mathbb{K})(\forall \xi_k \in \mathbb{R}) \quad |\phi_k'(\xi_k)|^2 = |\phi_k'(\xi_k) - \phi_k'(0)|^2 \leqslant \beta_k^2|\xi_k - 0|^2 \leqslant \beta^2|\xi_k|^2, \tag{2.40}$$

which yields

$$\big(\forall \mathsf{x} \in \ell^2(\mathbb{K})\big) \quad \sum_{k\in\mathbb{K}} |\phi_k'(\xi_k)|^2 \leqslant \beta^2 \sum_{k\in\mathbb{K}} |\xi_k|^2 = \beta^2\|\mathsf{x}\|^2 < +\infty \tag{2.41}$$

and allows us to conclude that φ is differentiable with $(\forall \mathsf{x} \in \ell^2(\mathbb{K}))$ $\nabla\varphi(\mathsf{x}) = (\phi_k'(\xi_k))_{k\in\mathbb{K}}$. Moreover,

$$\begin{aligned}\big(\forall \mathsf{x} \in \ell^2(\mathbb{K})\big)\big(\forall \mathsf{y} \in \ell^2(\mathbb{K})\big) \quad \|\nabla\varphi(\mathsf{x}) - \nabla\varphi(\mathsf{y})\|^2 &= \sum_{k\in\mathbb{K}} |\phi_k'(\xi_k) - \phi_k'(\eta_k)|^2 \\ &\leqslant \sum_{k\in\mathbb{K}} \beta_k^2|\xi_k - \eta_k|^2 \\ &\leqslant \beta^2\|\mathsf{x} - \mathsf{y}\|^2. \end{aligned} \tag{2.42}$$

Altogether, $\varphi\colon \ell^2(\mathbb{K}) \to \mathbb{R}$ is convex and differentiable with a β-Lipschitzian gradient. Hence $h = \varphi \circ F\colon \mathcal{H} \to \mathbb{R}$ is convex and differentiable, with $\nabla h = F^* \circ \nabla\varphi \circ F$. Furthermore, since F and F^* are isometries,

$$\begin{aligned}(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|\nabla h(x) - \nabla h(y)\| &= \|F^*(\nabla\varphi(Fx)) - F^*(\nabla\varphi(Fy))\| \\ &= \|\nabla\varphi(Fx) - \nabla\varphi(Fy)\| \\ &\leqslant \beta\|Fx - Fy\| \\ &= \beta\|x - y\|, \end{aligned} \tag{2.43}$$

which shows that ∇h is β-Lipschitzian. ∎

Remark 2.12. It is clear from the proof of Example 2.11 that if the condition

$$\phi_k \geqslant \phi_k(0) = 0 \tag{2.44}$$

is not satisfied for a finite number of indices $k \in \mathbb{K}$, the results remain valid. In particular, if $\mathcal{H}$ is finite-dimensional, (2.44) is not required. An example of a smooth convex function $\phi_k\colon \mathbb{R} \to \mathbb{R}$ with an explicit proximity operator is $\phi_k = d^2_{C_k}/2$, where C_k is a nonempty closed interval in $\mathbb{R}$ (set $\phi = |\cdot|^2/2$ in Example 2.3). In this case, (2.44) holds if and only if $0 \in C_k$. If $C_k = [1, +\infty[$, ϕ_k is the squared hinge loss [57]; if $C_k = [-\varepsilon, \varepsilon]$ for some $\varepsilon \in]0, +\infty[$, ϕ_k is the smooth Vapnik insensitive loss [4] (see also Example 2.4). Specializing Examples 2.3, 2.5, and 2.6, as well as (2.55) to $\mathcal{H} = \mathbb{R}$ provides further examples of functions $\phi_k = \psi_k \circ d_{C_k}$ of interest. For instance, taking ψ_k to be the Huber function (2.19) and $C_k = [1, +\infty[$ yields the modified Huber function of [65].

Example 2.13. Let M be a strictly positive integer and let $\varnothing \neq \mathbb{K} \subset \mathbb{N}$. For every $i \in \{1, \ldots, M\}$, suppose that $\mathcal{H}_i$ is a separable real Hilbert space with orthonormal basis $(e_{i,k})_{k\in\mathbb{K}}$. Let $\beta \in]0, +\infty[$, and let $\phi\colon \mathbb{R} \to \mathbb{R}$ be an even differentiable convex function such that $\phi \geqslant \phi(0) = 0$ and ϕ' is β-Lipschitzian. For every $(x_1, \ldots, x_M)$ in $\mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_M$, set $h(x_1, \ldots, x_M) = \sum_{k\in\mathbb{K}} \phi(\|(\langle x_1 \mid e_{1,k}\rangle, \ldots, \langle x_M \mid e_{M,k}\rangle)\|_2)$. Let $\gamma \in]0, +\infty[$, let $(x_1, \ldots, x_M) \in \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_M$, and set, for every $k \in \mathbb{K}$,

$$\alpha_k = \begin{cases} \dfrac{\phi'(\|(\langle x_1 \mid e_{1,k}\rangle, \ldots, \langle x_M \mid e_{M,k}\rangle)\|_2)}{\|(\langle x_1 \mid e_{1,k}\rangle, \ldots, \langle x_M \mid e_{M,k}\rangle)\|_2}, & \text{if } \max\limits_{1\leqslant i\leqslant M} |\langle x_i \mid e_{i,k}\rangle| > 0; \\ 0, & \text{otherwise,} \end{cases} \tag{2.45}$$

and

$$\delta_k = \begin{cases} \dfrac{\operatorname{prox}_{\gamma\phi}(\|(\langle x_1 \mid e_{1,k}\rangle, \ldots, \langle x_M \mid e_{M,k}\rangle)\|_2)}{\|(\langle x_1 \mid e_{1,k}\rangle, \ldots, \langle x_M \mid e_{M,k}\rangle)\|_2}, & \text{if } \max\limits_{1\leqslant i\leqslant M} |\langle x_i \mid e_{i,k}\rangle| > 0; \\ 1, & \text{otherwise.} \end{cases} \tag{2.46}$$

Then $h\colon \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_M \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x_1, \ldots, x_M) = \left(\sum_{k\in\mathbb{K}} \alpha_k \langle x_1 \mid e_{1,k}\rangle e_{1,k}, \ldots, \sum_{k\in\mathbb{K}} \alpha_k \langle x_M \mid e_{M,k}\rangle e_{M,k} \right), \tag{2.47}$$

and

$$\operatorname{prox}_{\gamma h}(x_1,\dots,x_M)=\bigg(\sum_{k\in\mathbb{K}}\delta_k\langle x_1\mid e_{1,k}\rangle e_{1,k},\dots,\sum_{k\in\mathbb{K}}\delta_k\langle x_M\mid e_{M,k}\rangle e_{M,k}\bigg). \tag{2.48}$$

Proof. Arguing as in (2.38)–(2.39), we obtain that h is real-valued and, arguing as in (2.40), we obtain $\sup_{k\in\mathbb{K}}\alpha_k\leqslant\beta$. Hence $(\forall i\in\{1,\dots,M\})\ (\alpha_k\langle x_i\mid e_{i,k}\rangle)_{k\in\mathbb{K}}\in\ell^2(\mathbb{K})$ and $\sum_{k\in\mathbb{K}}\alpha_k\langle x_i\mid e_{i,k}\rangle e_{i,k}\in\mathcal{H}_i$. Likewise, since ϕ is even, $\phi\geqslant\phi(0)=0$ [9, Proposition 11.7(i)], hence $\operatorname{prox}_{\gamma\phi}0=0$, and thus $\sup_{k\in\mathbb{K}}\delta_k\leqslant 1$. It therefore follows that $(\forall i\in\{1,\dots,M\})$ $(\delta_k\langle x_i\mid e_{i,k}\rangle)_{k\in\mathbb{K}}\in\ell^2(\mathbb{K})$ and $\sum_{k\in\mathbb{K}}\delta_k\langle x_i\mid e_{i,k}\rangle e_{i,k}\in\mathcal{H}_i$. Now let $\ell^2(\mathbb{K};\mathbb{R}^M)=\bigoplus_{k\in\mathbb{K}}\mathbb{R}^M$ be the space of square summable sequences with entries in $\mathbb{R}^M$, and set

$$\begin{cases}\varphi\colon\mathbb{R}^M\to\mathbb{R}\colon(\xi_1,\dots,\xi_M)\mapsto\phi(\|(\xi_1,\dots,\xi_M)\|_2)\\ U\colon\mathcal{H}_1\oplus\cdots\oplus\mathcal{H}_M\to\ell^2(\mathbb{K};\mathbb{R}^M)\colon(x_1,\dots,x_M)\mapsto\big((\langle x_1\mid e_{1,k}\rangle,\dots,\langle x_M\mid e_{M,k}\rangle)\big)_{k\in\mathbb{K}}.\end{cases} \tag{2.49}$$

Then U is a bijective isometry,

$$\begin{aligned}U^{-1}=U^*\colon\quad \ell^2(\mathbb{K};\mathbb{R}^M)&\to\ \mathcal{H}_1\oplus\cdots\oplus\mathcal{H}_M\\ \big((\mu_{1,k},\dots,\mu_{M,k})\big)_{k\in\mathbb{K}}&\mapsto\ \bigg(\sum_{k\in\mathbb{K}}\mu_{1,k}e_{1,k},\dots,\sum_{k\in\mathbb{K}}\mu_{M,k}e_{M,k}\bigg),\end{aligned} \tag{2.50}$$

and $h=\big(\bigoplus_{k\in\mathbb{K}}\varphi\big)\circ U$. In turn,

$$\begin{aligned}\nabla h&=U^*\circ\nabla\bigg(\bigoplus_{k\in\mathbb{K}}\varphi\bigg)\circ U\\ &=U^*\circ\nabla\bigg(\bigoplus_{k\in\mathbb{K}}\big(\phi\circ\|\cdot\|_2\big)\bigg)\circ U\\ &=U^*\circ\bigg(\bigotimes_{k\in\mathbb{K}}\nabla\big(\phi\circ\|\cdot\|_2\big)\bigg)\circ U,\end{aligned} \tag{2.51}$$

which yields (2.47) by applying Example 2.3 with $C=\{0\}$. On the other hand, using [9, Proposition 24.14], we obtain

$$\operatorname{prox}_{\gamma h}=\operatorname{prox}_{\gamma(\bigoplus_{k\in\mathbb{K}}\varphi)\circ U}=\big(U^*\circ\operatorname{prox}_{\gamma\bigoplus_{k\in\mathbb{K}}\varphi}\circ U\big)=U^*\circ\bigg(\bigotimes_{k\in\mathbb{K}}\operatorname{prox}_{\gamma\phi\circ\|\cdot\|_2}\bigg)\circ U, \tag{2.52}$$

which yields (2.48) by applying Example 2.3 with $C=\{0\}$. ∎

2.4. Function involving explicit infimal convolutions. If h is explicitly constructed in terms of a function $g\in\Gamma_0(\mathcal{H})$ as $h=g\,\square\,(\beta q)$ for some $\beta\in\left]0,+\infty\right[$, then, as in (2.1), we obtain

$$\nabla h=\beta\big(\operatorname{Id}-\operatorname{prox}_{g/\beta}\big)\quad\text{and}\quad\operatorname{prox}_{\gamma h}x=\operatorname{Id}+\frac{\gamma\beta}{\gamma\beta+1}\big(\operatorname{prox}_{(\gamma\beta+1)g/\beta}-\operatorname{Id}\big). \tag{2.53}$$

In the same spirit, we have the following construction.

Example 2.14. Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\beta \in]0,+\infty[$, and define $h = \beta q - (\beta q) \,\square\, \varphi$. Let $\gamma \in]0,+\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \beta \mathrm{prox}_{\varphi/\beta} x, \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = x - \beta\gamma \mathrm{prox}_{\frac{\varphi}{\beta(1+\beta\gamma)}}\left(\frac{1}{1+\beta\gamma} x\right). \tag{2.54}$$

Proof. It follows from [9, Proposition 12.30] that $\nabla h = \beta(\mathrm{Id} - (\mathrm{Id} - \mathrm{prox}_{\varphi/\beta})) = \beta \mathrm{prox}_{\varphi/\beta}$, which is β-Lipschitzian since $\mathrm{prox}_{\varphi/\beta}$ is nonexpansive [9, Proposition 12.28]. Finally the expression of $\mathrm{prox}_{\gamma h} x$ follows from [9, Proposition 24.8(viii)]. ∎

Remark 2.15. Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\beta \in]0,+\infty[$, let $\gamma \in]0,+\infty[$, and let $x \in \mathcal{H}$. Set $\varphi = \iota_C$ in Example 2.14. Then we derive that $h = \beta(q - d_C^2/2)\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient, and that

$$\nabla h(x) = \beta \mathrm{proj}_C x \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = x - \beta\gamma \mathrm{proj}_C\left(\frac{1}{1+\beta\gamma} x\right). \tag{2.55}$$

For $\beta = 1$, h is the generalized Huber function of [25, Example 3.2], that is,

$$(\forall x \in \mathcal{H}) \quad h(x) = \begin{cases} \langle x \mid \mathrm{proj}_C x\rangle - \dfrac{\|\mathrm{proj}_C x\|^2}{2}, & \text{if } x \notin C; \\ \dfrac{\|x\|^2}{2}, & \text{if } x \in C. \end{cases} \tag{2.56}$$

If $\mathcal{H} = \mathbb{R}$ and $C = [-\rho, \rho]$ for some $\rho \in]0,+\infty[$, h reduces to the standard Huber function (2.19).

3. Splitting algorithms. In this section, we review several proximal splitting algorithms which are relevant to our discussion and will be used in the numerical experiments (see [9] for the supporting theory). We start with the forward-backward splitting algorithm.

Algorithm 3.1 (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 β-Lipschitzian gradient, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 2/\beta[$ such that $0 < \inf_{n\in\mathbb{N}} \gamma_n \leqslant \sup_{n\in\mathbb{N}} \gamma_n < 2/\beta$. Let $x_0 \in \mathcal{H}$ and iterate*

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma_n \nabla h(x_n) \\ x_{n+1} = \mathrm{prox}_{\gamma_n f} y_n. \end{array} \right. \end{array} \tag{3.1}$$

Proposition 3.2. [33] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.1 and suppose that $\mathrm{Argmin}(f+h) \neq \varnothing$. Then there exists $x \in \mathrm{Argmin}(f+h)$ such that $x_n \rightharpoonup x$. In addition $(f+h)(x_n) - \inf(f+h)(\mathcal{H}) = o(1/n)$.*

Inertial variants of the above method have been popularized by [11]. They require additional storage capabilities but have been shown to be advantageous in terms of convergence speed in certain situations. The following implementation proposed in [18] guarantees convergence of the iterates.

Algorithm 3.3 (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 β-Lipschitzian gradient, let $\gamma \in]0, 1/\beta]$, and let $\alpha \in]2,+\infty[$. Let $x_0 = x_{-1} \in \mathcal{H}$ and iterate*

$$
\begin{aligned}
&\text{for } n = 0, 1, \ldots\\
&\left\lfloor
\begin{aligned}
z_n &= x_n + \frac{n-1}{n+\alpha}(x_n - x_{n-1})\\
y_n &= z_n - \gamma \nabla h(z_n)\\
x_{n+1} &= \operatorname{prox}_{\gamma f} y_n.
\end{aligned}
\right.
\end{aligned}
\tag{3.2}
$$

Proposition 3.4. [18] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.3 and suppose that $\operatorname{Argmin}(f+h) \neq \varnothing$. Then there exists $x \in \operatorname{Argmin}(f+h)$ such that $x_n \rightharpoonup x$. In addition $(f+h)(x_n) - \inf(f+h)(\mathcal{H}) = O(1/n^2)$.*

The next algorithm does not require smoothness of any of the functions.

Algorithm 3.5 (Douglas-Rachford). *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $0 \in \operatorname{sri}(\operatorname{dom} g - \operatorname{dom} f)$, 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*

$$
\begin{aligned}
&\text{for } n = 0, 1, \ldots\\
&\left\lfloor
\begin{aligned}
z_n &= \operatorname{prox}_{\gamma g} y_n\\
x_n &= \operatorname{prox}_{\gamma f}(2z_n - y_n)\\
y_{n+1} &= y_n + \lambda_n(x_n - z_n).
\end{aligned}
\right.
\end{aligned}
\tag{3.3}
$$

Proposition 3.6. [30] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.5 and suppose that $\operatorname{Argmin}(f+g) \neq \varnothing$. Then there exists $x \in \operatorname{Argmin}(f+g)$ such that $x_n \rightharpoonup x$.*

Although this feature will not be used in Section 4, it should be noted that the forward-backward [33], inertial forward-backward [7], and Douglas-Rachford [24] algorithms tolerate errors in the implementations of the proximity operators. The next three algorithms are specifically tailored to handle Problem 1.1. Although they also compute dual solutions, for brevity, we present only the primal convergence result for the error-free, unrelaxed formulations of these algorithms. The first one is known as the primal-dual forward-backward-forward algorithm; see [32] for details and [61] for a variable metric version.

Algorithm 3.7. *Consider the setting of Problem 1.1. 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*

(3.4)
$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
y_{1,n} = x_n - \gamma_n \Big(\sum_{i \in I} L_i^* v_{i,n}^* + \sum_{j \in J} L_j^* \big(\nabla h_j (L_j x_n) \big) \Big) \\
p_{1,n} = \operatorname{prox}_{\gamma_n f} y_{1,n} \\
\text{for every } i \in I \\
\left\lfloor \begin{array}{l}
y_{2,i,n} = v_{i,n}^* + \gamma_n L_i x_n \\
p_{2,i,n} = y_{2,i,n} - \gamma_n \operatorname{prox}_{g_i/\gamma_n} (y_{2,i,n}/\gamma_n) \\
q_{2,i,n} = p_{2,i,n} + \gamma_n L_i p_{1,n} \\
v_{i,n+1}^* = v_{i,n}^* - y_{2,i,n} + q_{2,i,n}
\end{array} \right. \\
q_{1,n} = p_{1,n} - \gamma_n \Big(\sum_{i \in I} L_i^* p_{2,i,n} + \sum_{j \in J} L_j^* \big(\nabla h_j (L_j p_{1,n}) \big) \Big) \\
x_{n+1} = x_n - y_{1,n} + q_{1,n}.
\end{array} \right.
\end{array}
$$

Proposition 3.8. [32] *Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 3.7. Then there exists a solution x to (1.6) such that $x_n \rightharpoonup x$.*

The following algorithm is an implementation of the forward-backward algorithm in a renormed primal-dual space (see [19, 27, 36, 40, 60] for special cases and variants, and [34] for a more general variable metric version).

Algorithm 3.9. *Consider the setting of Problem 1.1 and let $(\tau_n)_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $(\forall n \in \mathbb{N})\ \tau_{n+1} \geqslant \tau_n$. For every $i \in I$, let $v_{i,0}^* \in \mathcal{G}_i$, let $(\sigma_{i,n})_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $(\forall n \in \mathbb{N})\ \sigma_{i,n+1} \geqslant \sigma_{i,n}$. Suppose that*

(3.5)
$$
\sup_{n \in \mathbb{N}} \left(\sqrt{\tau_n \sum_{i \in I} \sigma_{i,n} \|L_i\|^2} + \frac{1}{2} \max \Big\{ \tau_n, \max_{i \in I} \sigma_{i,n} \Big\} \sum_{j \in J} \mu_j \|L_j\|^2 \right) < 1.
$$

Let $x_0 \in \mathcal{H}$ and iterate

(3.6)
$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
y_{1,n} = x_n - \tau_n \Big(\sum_{i \in I} L_i^* v_{i,n}^* + \sum_{j \in J} L_j^* \big(\nabla h_j (L_j x_n) \big) \Big) \\
x_{n+1} = \operatorname{prox}_{\tau_n f} y_{1,n} \\
z_n = 2 x_{n+1} - x_n \\
\text{for every } i \in I \\
\left\lfloor \begin{array}{l}
y_{2,i,n} = v_{i,n}^* + \sigma_{i,n} (L_i z_n) \\
v_{i,n+1}^* = y_{2,i,n} - \sigma_{i,n} \operatorname{prox}_{g_i/\sigma_{i,n}} (y_{2,i,n}/\sigma_{i,n}).
\end{array} \right.
\end{array} \right.
\end{array}
$$

Proposition 3.10. [34] *Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 3.9. Then there exists a solution x to (1.6) such that $x_n \rightharpoonup x$.*

The next algorithm, which was first proposed in [2] in the case when $J = \varnothing$, was extended in [29] to a block-coordinate and block-iterative asynchronous method. The following version, which explicitly exploits smooth functions, is proposed in [43].

Algorithm 3.11. *Consider the setting of Problem 1.1 and let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $0 < \inf_{n \in \mathbb{N}} \gamma_n \leqslant \sup_{n \in \mathbb{N}} \gamma_n < +\infty$. For every $k \in I \cup J$, let $v_{k,0}^* \in \mathcal{G}_k$, let $(\mu_{k,n})_{n \in \mathbb{N}}$*

be a sequence in $]0,+\infty[$ *such that* $0 < \inf_{n\in\mathbb{N}} \mu_{k,n} \leqslant \sup_{n\in\mathbb{N}} \mu_{k,n} < +\infty$*, and let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,2[$ *such that* $0 < \inf_{n\in\mathbb{N}} \lambda_n \leqslant \sup_{n\in\mathbb{N}} \lambda_n < 2$*. Let* $x_0 \in \mathcal{H}$ *and iterate*

(3.7)
for $n = 0, 1, \ldots$
  $l_n^* = \sum_{k\in I\cup J} L_k^* v_{k,n}^*$
  $a_n = \operatorname{prox}_{\gamma_n f}(x_n - \gamma_n l_n^*)$
  $a_n^* = \gamma_n^{-1}(x_n - a_n) - l_n^*$
  for $i \in I$
    $l_{i,n} = L_i x_n$
    $b_{i,n} = \operatorname{prox}_{\mu_{i,n} g_i}\big(l_{i,n} + \mu_{i,n} v_{i,n}^*\big)$
    $b_{i,n}^* = v_{i,n}^* + \mu_{i,n}^{-1}(l_{i,n} - b_{i,n})$
    $t_{i,n} = b_{i,n} - L_i a_n$
  for $j \in J$
    $l_{j,n} = L_j x_n$
    $b_{j,n} = l_{j,n} - \mu_{j,n}\big(\nabla h_j(l_{j,n}) - v_{j,n}^*\big)$
    $b_{j,n}^* = \nabla h_j(b_{j,n})$
    $t_{j,n} = b_{j,n} - L_j a_n$
  $t_n^* = a_n^* + \sum_{k\in I\cup J} L_k^* b_{k,n}^*$
  $\tau_n = \|t_n^*\|^2 + \sum_{k\in I\cup J} \|t_{k,n}\|^2$
  if $\tau_n = 0$
    $x = a_n$
    terminate.
  if $\tau_n > 0$
    $\theta_n = \dfrac{\lambda_n}{\tau_n} \max\Big\{0, \langle x_n \mid t_n^*\rangle - \langle a_n \mid a_n^*\rangle + \sum_{k\in I\cup J}\big(\langle t_{k,n} \mid v_{k,n}^*\rangle - \langle b_{k,n} \mid b_{k,n}^*\rangle\big)\Big\}$
    $x_{n+1} = x_n - \theta_n t_n^*$
    for $k \in I \cup J$
      $v_{k,n+1}^* = v_{k,n}^* - \theta_n t_{k,n}$.

Proposition 3.12. [43] *Either Algorithm 3.11 terminates at a solution* x *to* (1.6) *in a finite number of iterations, or it generates an infinite sequence* $(x_n)_{n\in\mathbb{N}}$ *which converges weakly to a solution to* (1.6).

4. Applications and numerical illustrations. We illustrate the viewpoint formulated in the Introduction, which suggests that it may be computationally advantageous to activate smooth functions proximally in certain instances.

We compare the splitting methods reviewed in Section 3 on various digital image restoration and reconstruction problems. All images have $\sqrt{N} \times \sqrt{N}$ pixels and therefore the underlying Hilbert space is $\mathcal{H} = \mathbb{R}^N$ ($N \in \{96^2, 128^2, 512^2\}$) equipped with the standard Euclidean norm $\|\cdot\|_2$. All the algorithms guarantee the convergence of their iterates $(x_n)_{n\in\mathbb{N}}$ to a solution x of the underlying optimization problem. Let us also note that this set of experiments constitutes the first implementation of Algorithm 3.11 to image recovery. The simulations are run in Python on a personal computer running Linux Ubuntu version 18.04 with a 2.60GHz dual-core processor and 8GB of RAM. Finally, the normalization used to plot the decibel value

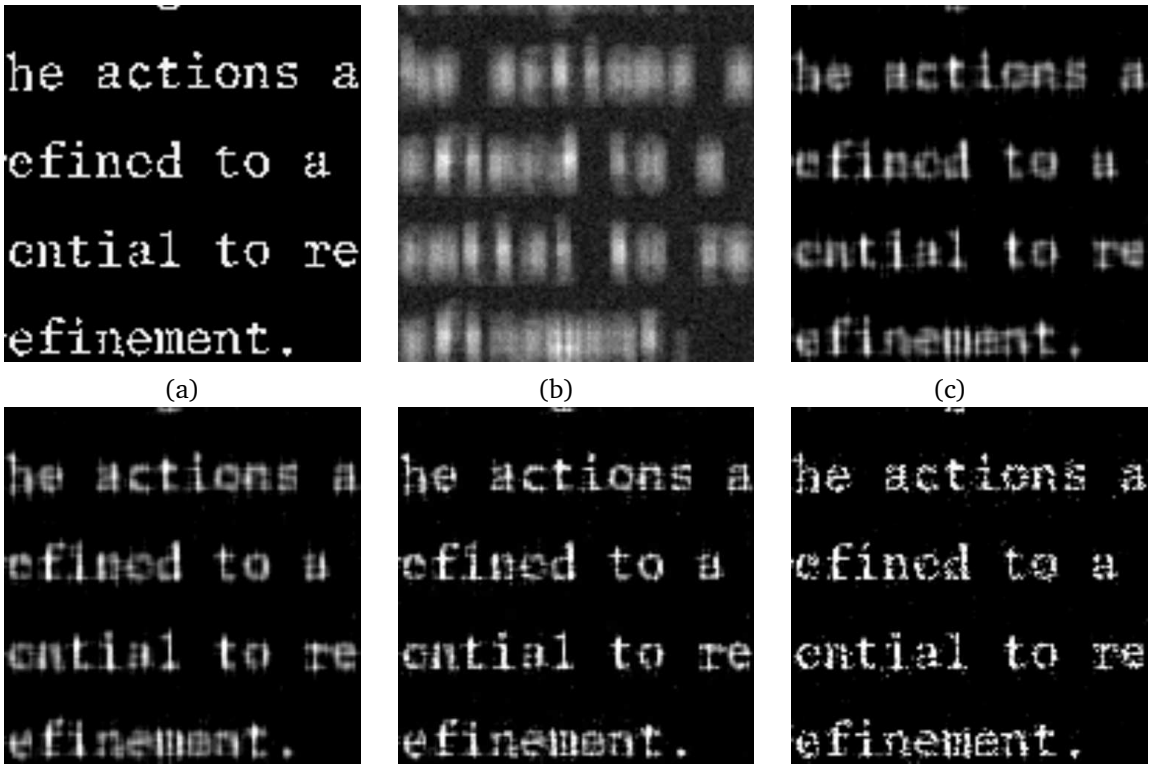

Figure 4.1. *(a) Original image $\overline{x}$. (b) Degraded image y. (c) Image restored by the forward-backward algorithm (Algorithm 3.1) after 50 iterations. (d) Image restored by the inertial forward-backward algorithm (Algorithm 3.3) after 50 iterations. (e) Image restored by the Douglas-Rachford algorithm (Algorithm 3.5) after 50 iterations. (f) Restored image (all algorithms yield visually equivalent images).*

of the squared distance of an iterate x_n to a solution x is

$$20 \log_{10} \frac{\|x_n - x\|_2}{\|x_0 - x\|_2}, \tag{4.1}$$

and that used for the objective value $\varphi(x_n)$ at iteration n is

$$10 \log_{10} \frac{\varphi(x_n) - \varphi(x)}{\varphi(x_0) - \varphi(x)}. \tag{4.2}$$

Remark 4.1. The applications to be considered below are instances of Problem 1.1 in which f has bounded domain and the remaining functions have full domain, which ensures existence of at least one solution [9, Corollary 11.16(i)]. In addition, the inclusion (1.5) holds by virtue of [32, Proposition 4.3(ii)].

4.1. Sparse image deconvolution. We consider a very basic instance of Problem 1.1 with only two functions. More specifically, we compare the numerical behavior of the forward-backward algorithms of Propositions 3.2 and 3.4 with that of the Douglas-Rachford algorithm

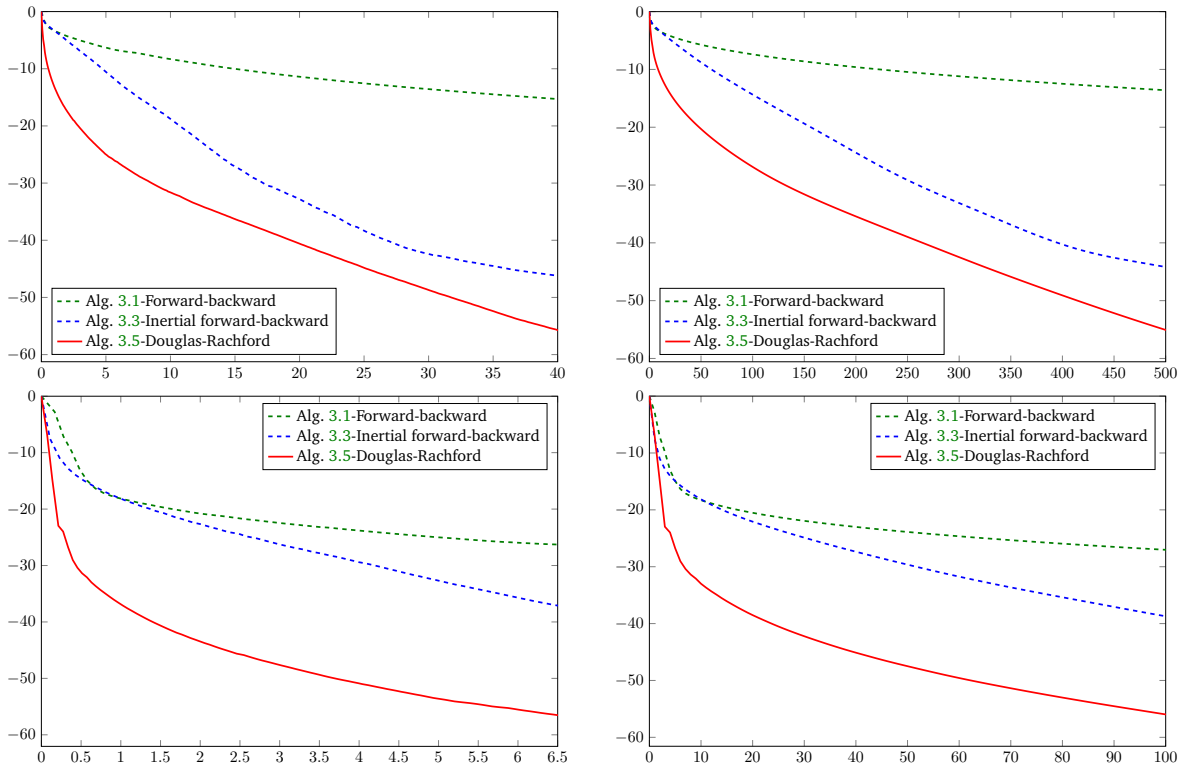

Figure 4.2. *Top left: Normalized distance in dB to the asymptotic image produced by each algorithm versus execution time in seconds. Top right: Normalized distance in dB to the asymptotic image produced by each algorithm versus iteration number. In this experiment the objective function values remain finite and can therefore be displayed. Bottom left: Normalized objective function of* (4.3) *in dB versus execution time in seconds. Bottom right: Normalized objective function of* (4.3) *in dB versus iteration number.*

of Proposition 3.6, which is a fully proximal method. Note that an elementary comparison of these three algorithms to minimize $f + h$ already appears in Fig. 1.1, where $f = 0$ and $h = \varphi$. The images have size 128×128.

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 15×5 and w is a Gaussian white noise realization (see Fig. 4.1(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 to

$$\underset{x \in C}{\text{minimize}} \;\; \|x\|_1 + \frac{1}{2}\|Hx - y\|_2^2. \tag{4.3}$$

Now set $f = \|\cdot\|_1 + \iota_C$, set $h = \|H \cdot -y\|_2^2/2$, and let $\gamma \in \left]0, +\infty\right[$. Then $f \in \Gamma_0(\mathcal{H})$ and $\operatorname{prox}_{\gamma f} = \operatorname{proj}_C \circ \operatorname{soft}_\gamma$ [9, Propositions 24.12(ii) and 24.47], where $\operatorname{soft}_\gamma$ is the soft thresholder on $[-\gamma, \gamma]$; the projector proj_C is implemented by setting to 0 the pixel values less than 0, and to 255 those larger than 255. On the other hand, h is smooth, with gradient and

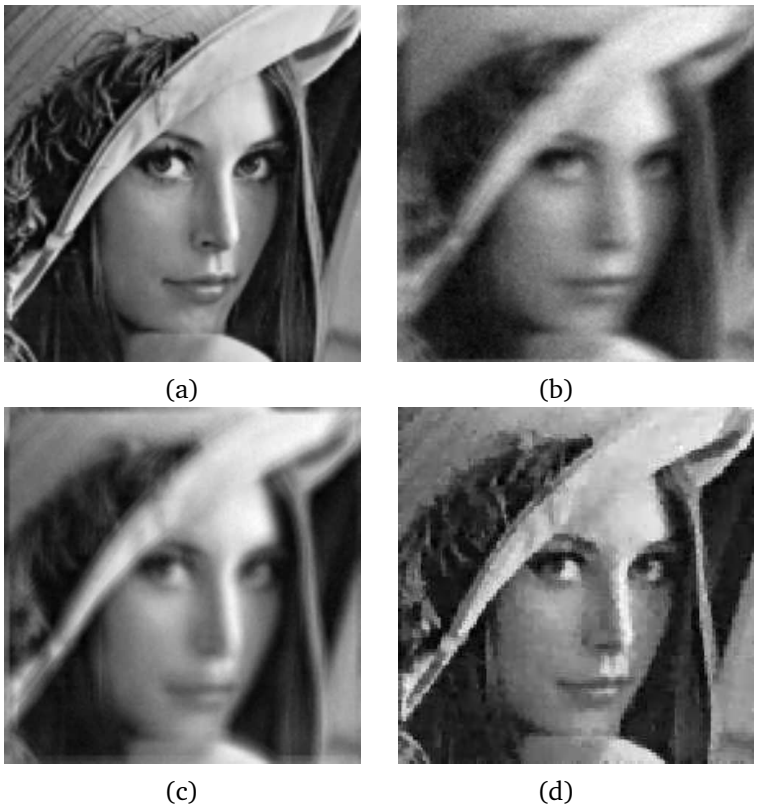

Figure 4.3. *(a) Original image $\overline{x}$. (b) Degraded image y_1. (c) Degraded image y_2. (d) Reconstructed image (all algorithms yield visually equivalent images).*

proximity operator provided by Example 2.2 as

$$\nabla h\colon x \mapsto H^*(Hx - y) \quad \text{and} \quad \operatorname{prox}_{\gamma h}\colon x \mapsto (\mathrm{Id} + \gamma H^* H)^{-1}(x + \gamma H^* y). \tag{4.4}$$

The linear operator H models a convolution and it is representable by a block-circulant matrix. The computation of the inverse in (4.4) is therefore straightforward via the fast Fourier transform [3].

We solve (4.3) with the forward-backward algorithms (Algorithm 3.1 and Algorithm 3.3), as well as with the fully proximal Douglas-Rachford algorithm (Algorithm 3.5). The algorithms are initialized at zero and implemented with the parameters for which they seem to perform best, that is, $\gamma_n \equiv 1.99/\beta$ for Algorithm 3.1, $\gamma = 1/\beta$ and $\alpha = 3$ for Algorithm 3.3, and $\gamma = 30$ and $\lambda_n \equiv 1.9$ for Algorithm 3.5. The results of Figs. 4.1–4.2 show a superior performance for Algorithm 3.5, which is fully proximal.

4.2. Multiview image reconstruction from partial diffraction data. We consider the problem of reconstructing a 128×128 image $\overline{x}$ from a partial observation of its diffraction over some frequency range R, possibly with measurement errors [56]. To exploit this information we use the soft constraint penalty d_E associated with the set

$$E = \big\{x \in \mathbb{R}^N \;\big|\; (\forall k \in R)\ \widehat{x}(k) = \widehat{\overline{x}}(k)\big\}, \tag{4.5}$$

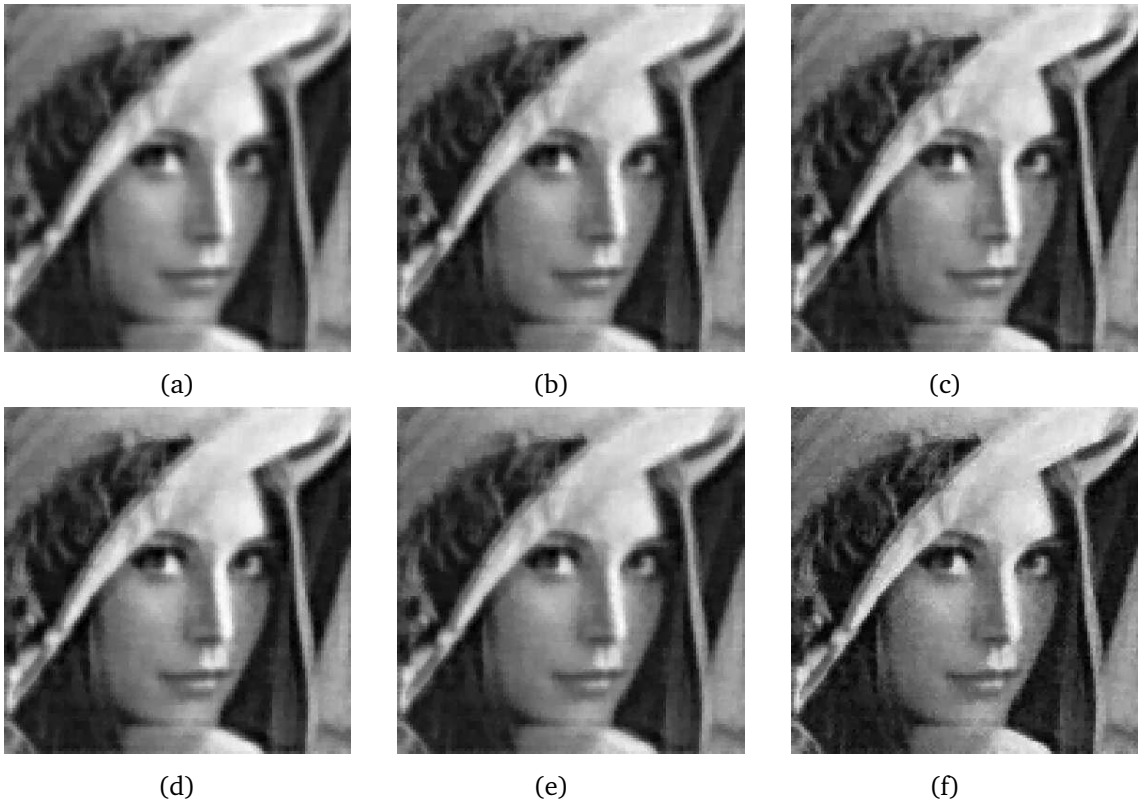

Figure 4.4. *(a) Image restored by Pb. 4.2/Alg. 3.7-S after 20 iterations. (b) Image restored by Pb. 4.2/Alg. 3.9-S after 20 iterations. (c) Image restored by Pb. 4.2/Alg. 3.11-S after 20 iterations. (d) Image restored by Pb. 4.3/Alg. 3.7-P after 20 iterations. (e) Image restored by Pb. 4.3/Alg. 3.9-P after 20 iterations. (f) Image restored by Pb. 4.3/Alg. 3.11-P after 20 iterations.*

where $\widehat{x}$ denotes the two-dimensional discrete Fourier transform of x. The set R contains the frequencies in $\{0,\dots,15\}^2$ as well as those resulting from the symmetry properties of the discrete Fourier transform. In addition, we have at our disposal two blurred noisy observations of $\overline{x}$, namely (see Fig. 4.3(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 convolutional blurs with kernels of sizes 3×11 and 7×5, 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. We use $C = [0,255]^N$ as a hard constraint set. Finally, we utilize a discrete version of the Gauss-TV penalty of Remark 2.10 to control oscillations in the reconstructed image. This leads to the formulation

$$\underset{x\in C}{\text{minimize}} \;\; \frac{1}{2}d_E(x) + \frac{2}{5}h(Dx) + \frac{3}{4}\|H_1x - y_1\|_2^2 + \frac{3}{4}\|H_2x - y_2\|_2^2, \tag{4.6}$$

where $D\colon \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N\colon x \mapsto (G_1x, G_2x)$, G_1 and G_2 being horizontal and vertical discrete difference operators, and where $(\forall (y_1,y_2) = ((\eta_{1,k})_{1\leqslant k\leqslant N}, (\eta_{2,k})_{1\leqslant k\leqslant N}) \in \mathcal{G}_2 = \mathbb{R}^N \times \mathbb{R}^N)$ $h(y_1,y_2) = \sum_{k=1}^N \mathfrak{h}_2(\|(\eta_{1,k},\eta_{2,k})\|_2)$, $\mathfrak{h}_2$ being the Huber function (2.19). We derive from (4.6) two versions of Problem 1.1.

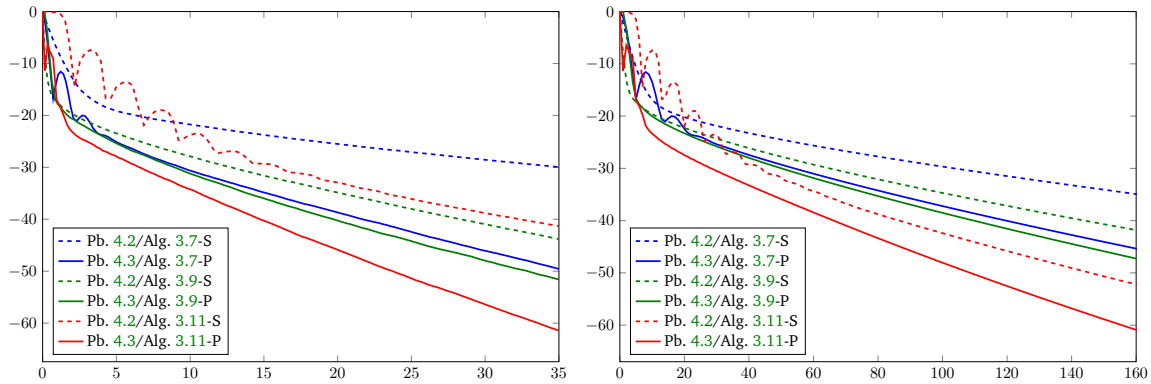

Figure 4.5. *Left: Normalized distance in dB to the asymptotic image produced by each algorithm versus execution time in seconds. Right: Normalized distance in dB to the asymptotic image produced by each algorithm versus iteration number. Each algorithm is represented by a given color; the solid line corresponds to the fully proximal implementation and the dashed line to the implementation with gradient steps for the smooth functions.*

Problem 4.2. In Problem 1.1, set $f = \iota_C$, $I = \{1\}$, $g_1 = 0.5d_E$, $L_1 = \mathrm{Id}$, $J = \{2,3,4\}$, $h_2 = 0.4h$, $L_2 = D$, $h_3 = 0.75\|\cdot - y_1\|_2^2$, $L_3 = H_1$, $h_4 = 0.75\|\cdot - y_2\|_2^2$, and $L_4 = H_2$.

Problem 4.3 (fully proximal). In Problem 1.1, set $f = \iota_C$, $I = \{1,2,3,4\}$, $J = \varnothing$, $g_1 = 0.5d_E$, $L_1 = \mathrm{Id}$, $g_2 = 0.4h$, $L_2 = D$, $g_3 = 0.75\|H_1 \cdot - y_1\|_2^2$, $L_3 = \mathrm{Id}$, $g_4 = 0.75\|H_2 \cdot - y_2\|_2^2$, and $L_4 = \mathrm{Id}$.

We apply to these two problems Algorithms 3.7, 3.9, and 3.11 with all initial vectors set to 0. The following parameters are used, where $\beta = \sqrt{\sum_{i\in I}\|L_i\|^2} + \sum_{j\in J}\mu_j\|L_j\|^2$ (these parameters were found to optimize the performance of each algorithm):

- Algorithm 3.7-S (with smooth terms for Problem 4.2): $\gamma_n \equiv 0.99/\beta$.
- Algorithm 3.7-P (fully proximal for Problem 4.3): $\gamma_n \equiv 0.99/\beta$.
- Algorithm 3.9-S (with smooth terms for Problem 4.2): $\sigma_{1,n} \equiv 8/(5\beta)$ and $\tau_n \equiv 8/(5\beta)$.
- Algorithm 3.9-P (fully proximal for Problem 4.3): $\sigma_{1,n} \equiv 1/(2\beta)$, $\sigma_{2,n} \equiv 1/(2\beta)$, $\sigma_{3,n} \equiv 3/\beta$, $\sigma_{4,n} \equiv 3/\beta$, and $\tau_n \equiv 1/\beta$.
- Algorithm 3.11-S (with smooth terms for Problem 4.2): $\gamma_n \equiv 0.4$, $\mu_{1,n} \equiv 1.0$, $\mu_{2,n} \equiv 2.49$, $\mu_{3,n} \equiv 0.65$, $\mu_{4,n} \equiv 0.65$, and $\lambda_n \equiv 1.99$.
- Algorithm 3.11-P (fully proximal for Problem 4.3): $\gamma_n \equiv 0.25$, $\mu_{1,n} \equiv 1.0$, $\mu_{2,n} \equiv 1.5$, $\mu_{3,n} \equiv 1.0$, $\mu_{4,n} \equiv 1.0$, and $\lambda_n \equiv 1.99$.

The proximity operators and the gradients used in these experiments follow from [9, Example 24.28] for g_1, Example 2.13 (with $M = 2$ and $\mathbb{K} = \{1,\dots,N\}$, and using [9, Example 24.9] to get the proximity operator of $\mathfrak{h}_2$) for h_2, and (4.4) for g_3 and g_4. On the other hand $\mathrm{prox}_{\gamma f} = \mathrm{proj}_C$. We have $\|D\|^2 = 8$ and $\|H_1\|^2 = \|H_2\|^2 = 1$. On the other hand, the functions h_2, h_3, and h_4 are differentiable with a Lipschitzian gradient, and their Lipschitz constants are respectively 0.4, 0.75, and 0.75. Thus, all the assumptions required by the algorithms are satisfied. The results shown in Figs. 4.4 and 4.5 illustrate the faster convergence of the iterates $(x_n)_{n\in\mathbb{N}}$ of fully proximal algorithms to a solution x compared to their gradient-based versions, both in terms of computation time and iterations. Note that these primal-dual algorithms do not guarantee Fejér monotonicity in the primal space, i.e., that

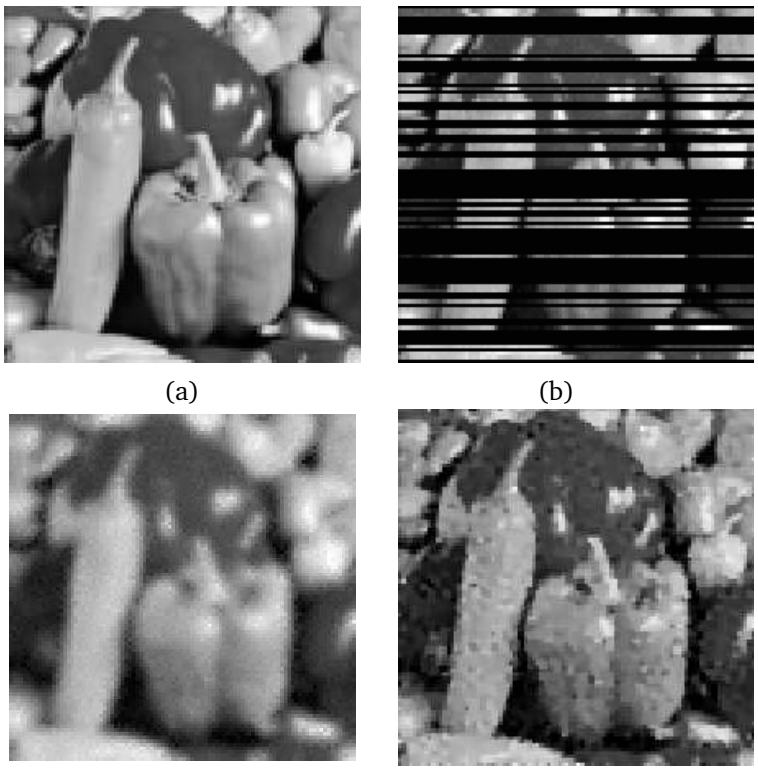

Figure 4.6. *(a) Original image $\overline{x}$. (b) Degraded image y_1. (c) Degraded image y_2. (d) Reconstructed image (all algorithms yield visually equivalent images).*

$(\|x_n - x\|)_{n\in\mathbb{N}}$ goes to 0 monotonically. This is confirmed by the oscillations seen in Fig. 4.5. Finally, Fig. 4.3(d) shows that the formulation (4.6) provides a faithful recovery of $\overline{x}$.

4.3. Image interpolation. We consider the problem of reconstructing the 96×96 original image $\overline{x}$ shown in Fig. 4.6(a) from a noisy occulted version y_1 and a blurred and noisy measurement y_2 (see Fig. 4.6(b)–(c)). The occulted version is missing 57 lines and the observed line numbers are indexed by $R \subset \{1, \dots, N\}$. The observations are $y_1 = M\overline{x} + w_1$ and $y_2 = H\overline{x} + w_2$, where M is the masking operator zeroing the lines not indexed by R and where H is a blurring operator. By contrast with the previous experiments, the blurring is nonstationary and therefore does not correspond to a convolution operation. More precisely, the action of the blurring operator H on a given pixel (i, j) is to replace it by an average of the neighboring pixels weighted by an isotropic Gaussian kernel centered at (i, j) and with random standard deviation $\sigma_{i,j} \in [0, 1]$. Finally $\|H\| = 1$, while w_1 and w_2 are realizations of Gaussian white noises such that the image-to-noise ratio for y_1 is 25.9 dB and the blurred image-to-noise ratio for y_2 is 31.0 dB. We denote by $(y_1^{(i)})_{i\in R}$ the nonzero lines of y_1 corresponding to the observed lines of $\overline{x}$.

To model this interpolation problem, we use $C = [0, 255]^N$ as a hard constraint as well as the total variation penalty. In addition, we fit the observed lines via the penalty $x \mapsto$

$\sum_{i\in R} 10\|x^{(i)} - y_1^{(i)}\|_2$ and the degraded image via the penalty $x \mapsto 5\|Hx - y_2\|_2^2$. This leads to the formulation

$$\underset{x\in C}{\text{minimize}} \ \|Dx\|_{1,2} + 10\sum_{i\in R}\|x^{(i)} - y_1^{(i)}\|_2 + 5\|Hx - y_2\|_2^2, \tag{4.7}$$

where D is as in (4.6) and $(\forall (y_1, y_2) \in \mathcal{G}_2 = \mathbb{R}^N \times \mathbb{R}^N)\ \|(y_1, y_2)\|_{1,2} = \sum_{k=1}^N \|(\eta_{1,k}, \eta_{2,k})\|_2$. Two versions of Problem 1.1 are employed.

Problem 4.4. In Problem 1.1, set $f = \iota_C$, $I = \{1, 2\}$, $J = \{3\}$, $g_1 = \|\cdot\|_{1,2}$, $L_1 = D$, $g_2 = 10\sum_{i\in R}\|x^{(i)} - y_1^{(i)}\|_2$, $L_2 = \mathrm{Id}$, $h_3 = 5\|\cdot - y_2\|_2^2$, and $L_3 = H$.

Problem 4.5 (fully proximal). In Problem 1.1, set $f = \iota_C$, $I = \{1, 2, 3\}$, $J = \varnothing$, $g_1 = \|\cdot\|_{1,2}$, $L_1 = D$, $g_2 = 10\sum_{i\in R}\|x^{(i)} - y_1^{(i)}\|_2$, $L_2 = \mathrm{Id}$, $g_3 = 5\|H\cdot - y_2\|_2^2$, and $L_3 = \mathrm{Id}$.

We apply to Problems 4.4 and 4.5 Algorithms 3.7, 3.9, and 3.11 with all initial vectors set to 0. The following parameters are used, where $\beta = \sqrt{\sum_{i\in I}\|L_i\|^2 + \sum_{j\in J}\mu_j\|L_j\|^2}$ (these parameters were found to optimize the performance of each algorithm):

- Algorithm 3.7-S (with smooth terms for Problem 4.4): $\gamma_n \equiv 0.99/\beta$.
- Algorithm 3.7-P (fully proximal for Problem 4.5): $\gamma_n \equiv 0.99/\beta$.
- Algorithm 3.9-S (with smooth terms for Problem 4.4): $\sigma_{1,n} \equiv 2/(5\beta)$ and $\tau_n \equiv 0.1/\beta$.
- Algorithm 3.9-P (fully proximal for Problem 4.5): $\sigma_{1,n} \equiv 1/\beta$, $\sigma_{2,n} \equiv 1/\beta$, $\sigma_{3,n} \equiv 1/\beta$, and $\tau_n \equiv 1/\beta$.
- Algorithm 3.11-S (with smooth terms for Problem 4.4): $\gamma_n \equiv 1$, $\mu_{1,n} \equiv 0.1$, $\mu_{2,n} \equiv 0.1$, $\mu_{3,n} \equiv 0.01$, and $\lambda_n \equiv 1.9$.
- Algorithm 3.11-P (fully proximal for Problem 4.5): $\gamma_n \equiv 0.5$, $\mu_{1,n} \equiv 1.0$, $\mu_{2,n} \equiv 0.1$, $\mu_{3,n} \equiv 0.01$, and $\lambda_n \equiv 1.9$.

The proximity operators of f and g_3 are discussed in Section 4.2. The proximity operator of g_1 is computed similarly to that of g_2 in Problem 4.3 since, in view of Lemma 2.1, we can apply (2.48) with $\phi = |\cdot|$. It follows from [9, Propositions 24.8(ii) and 24.11] and [9, Example 24.20] that the proximity operator of g_2 for index $\gamma \in]0, +\infty[$ is

$$\text{(4.8)}\quad \mathrm{prox}_{\gamma g_2}\colon (x^{(i)})_{1\leqslant i\leqslant N} \mapsto (p^{(i)})_{1\leqslant i\leqslant N}, \quad \text{where} \quad (\forall i \in \{1, \dots, N\})$$
$$p^{(i)} = \begin{cases} x^{(i)}, & \text{if } i \notin R; \\ y_1^{(i)} + \left(1 - \dfrac{\gamma}{\max\left\{\left\|x^{(i)} - y_1^{(i)}\right\|_2, \gamma\right\}}\right)\left(x^{(i)} - y_1^{(i)}\right), & \text{if } i \in R. \end{cases}$$

As seen in Fig. 4.6(d), the missing lines are satisfactorily reconstructed. On the other hand, the convergence profiles displayed in Fig. 4.7 indicate that the fully proximal algorithms behave better than their gradient-based counterparts.

4.4. Inconsistent convex feasibility problems.

4.4.1. Mathematical model. As mentioned in the Introduction, the convex feasibility formalism first proposed in [64] has enjoyed continued interest from the image recovery community [14, 23, 47, 58]. A structured formulation of this problem is the following.

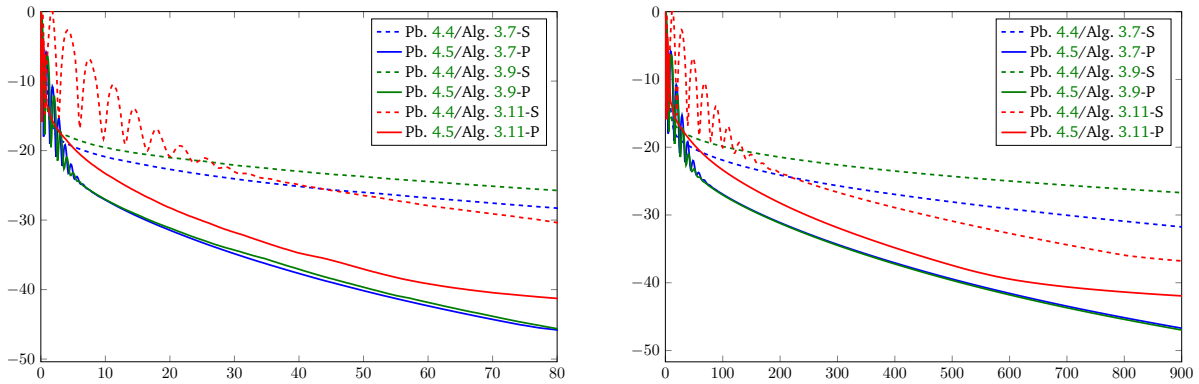

Figure 4.7. *Left: Normalized distance in dB to the asymptotic image produced by each algorithm versus execution time in seconds. Right: Normalized distance in dB to the asymptotic image produced by each algorithm versus iteration number.*

Problem 4.6. Let $\mathcal{H}$ be a real Hilbert space, let E be a nonempty closed convex subset of $\mathcal{H}$, let K be a nonempty finite subset of $\mathbb{N}$, and let $(\mathcal{G}_k)_{k\in K}$ be a family of real Hilbert spaces. For every $k \in K$, suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$ and let C_k be a nonempty closed convex subset of $\mathcal{G}_k$. The goal is to

$$\text{find} \quad x \in E \quad \text{such that} \quad (\forall k \in K) \quad L_k x \in C_k. \tag{4.9}$$

In applications, because of possible inaccuracies in *a priori* knowledge, unmodeled dynamics, or too aggressive confidence bounds on stochastic constraints, the above convex feasibility problem may turn out to be inconsistent [16, 22, 23, 39, 63], i.e., $E \cap \bigcap_{k\in K} L_k^{-1}(C_k) = \varnothing$. To deal with this situation, we propose the following variational formulation as a relaxation of (4.9).

Problem 4.7. Consider the setting of Problem 4.6. Let (I, J) be a partition of K such that, for every $i \in I$, $\phi_i \in \Gamma_0(\mathbb{R})$ is an even function that vanishes only at 0 and, for every $j \in J$, $\psi_j \colon \mathbb{R} \to \mathbb{R}$ is an even differentiable convex function that vanishes only at 0 with a Lipschitzian derivative. The problem is to

$$\operatorname*{minimize}_{x\in E} \quad \sum_{i\in I} \phi_i\big(d_{C_i}(L_i x)\big) + \sum_{j\in J} \psi_j\big(d_{C_j}(L_j x)\big). \tag{4.10}$$

Problem 4.7 unifies several formulations that have been proposed in the literature as surrogates to the possibly inconsistent Problem 4.6:

- If (4.9) happens to have solutions, they are the same as those of (4.10).
- In (4.10), E plays the role of a hard constraint; if no such constraint is present, one can set $E = \mathcal{H}$. Further hard constraints can be modeled by taking $\phi_i = \iota_{\{0\}}$ for certain $i \in I$.
- Suppose that $J = \varnothing$ and $(\forall i \in I)\ \phi_i = \iota_{\{0\}}$. Then (4.10) reverts to (4.9).
- Let $(\omega_j)_{j\in J}$ be real numbers in $]0, 1]$ such that $\sum_{j\in J} \omega_j = 1$. Suppose that $I = \varnothing$, $E = \mathcal{H}$, and $(\forall j \in J)\ \mathcal{G}_j = \mathcal{H}$, $L_j = \mathrm{Id}$, and $\psi_j = \omega_j |\cdot|^2/2$. Then we recover

the least-squares formulation of [22], namely the problem of minimizing $\sum_{j\in J}\omega_j d_{C_j}^2$ over $\mathcal{H}$.

- Let $(\omega_j)_{j\in J}$ be real numbers in $]0,1]$ such that $\sum_{j\in J}\omega_j = 1$. Suppose that $I = \varnothing$ and $(\forall j \in J)\ \mathcal{G}_j = \mathcal{H}$, $L_j = \mathrm{Id}$, and $\psi_j = \omega_j|\cdot|^2/2$. Then we recover the hard-constrained least-squares formulation of [26], namely

$$\operatorname*{minimize}_{x\in E}\ \frac{1}{2}\sum_{j\in J}\omega_j d_{C_j}^2(x). \tag{4.11}$$

  For $J = \{1\}$, this framework reduces to the formulation proposed in [39, 63].
- Let $(\omega_j)_{j\in J}$ be real numbers in $]0,1]$ such that $\sum_{j\in J}\omega_j = 1$. Suppose that $I = \varnothing$ and (J_1, J_2) is a partition of J. Suppose further that $(\forall j \in J_1)\ \mathcal{G}_j = \mathcal{H}$, $L_j = \mathrm{Id}$, and $\psi_j = \omega_j|\cdot|^2/2$, and that $(\forall j \in J_2)\ \psi_j = \omega_j|\cdot|^2/2$. Then we recover the formulation of [15], namely

$$\operatorname*{minimize}_{x\in E}\ \frac{1}{2}\sum_{j\in J_1}\omega_j d_{C_j}^2(x) + \frac{1}{2}\sum_{j\in J_2}\omega_j d_{C_j}^2(L_j x). \tag{4.12}$$

- Let $(\mathcal{H}_\ell)_{1\leqslant\ell\leqslant m}$ and $(\mathcal{K}_k)_{1\leqslant k\leqslant p}$ be real Hilbert spaces. For every $\ell \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, let C_ℓ be a nonempty closed convex subset of $\mathcal{H}_\ell$, let E_k be a nonempty closed convex subset of $\mathcal{K}_k$, and let $M_{k\ell} \in \mathcal{B}(\mathcal{H}_\ell, \mathcal{K}_k)$. Consider the multivariate convex feasibility problem

$$\text{find}\ \ x_1 \in C_1, \ldots, x_m \in C_m\ \ \text{such that}\ \ \sum_{\ell=1}^m M_{1\ell}x_\ell \in E_1, \ldots, \sum_{\ell=1}^m M_{p\ell}x_\ell \in E_p. \tag{4.13}$$

  Now let (I, J_0) be a partition of $\{1,\ldots,m\}$, let $(\phi_i)_{i\in I}$ be even functions in $\Gamma_0(\mathbb{R})$ vanishing only at 0, and, for every $j \in J_0$, let $\psi_j\colon \mathbb{R}\to\mathbb{R}$ be an even differentiable convex function that vanishes only at 0 with a Lipschitzian derivative. A relaxation of (4.13) proposed in [13, Section 3.3] is

$$\operatorname*{minimize}_{x_1\in\mathcal{H}_1,\ldots,x_m\in\mathcal{H}_m}\ \sum_{i\in I}\phi_i\big(d_{C_i}(x_i)\big) + \sum_{j\in J_0}\psi_j\big(d_{C_j}(x_j)\big) + \frac{1}{2p}\sum_{k=1}^p d_{E_k}^2\bigg(\sum_{\ell=1}^m M_{k\ell}x_\ell\bigg). \tag{4.14}$$

  Now let

$$\begin{cases} \mathcal{H} = \bigoplus_{\ell=1}^m \mathcal{H}_\ell,\ \mathcal{G}_0 = \bigoplus_{k=1}^p \mathcal{K}_k \\ L_0\colon \mathcal{H}\to\mathcal{G}_0\colon (x_1,\ldots,x_m) \mapsto \big(\sum_{\ell=1}^m M_{1\ell}x_\ell, \ldots, \sum_{\ell=1}^m M_{p\ell}x_\ell\big) \\ C_0 = E_1\times\cdots\times E_p,\ \psi_0 = |\cdot|^2/(2p) \\ (\forall k \in I\cup J_0)\ \mathcal{G}_k = \mathcal{H}_k,\ L_k\colon \mathcal{H}\to\mathcal{G}_k\colon (x_1,\ldots,x_m)\mapsto x_k \\ J = \{0\}\cup J_0,\ E = \mathcal{H}. \end{cases} \tag{4.15}$$

  Then (4.13) reduces to an instance of (4.9), and (4.14) of (4.10).

Remark 4.8. Problem 4.7 corresponds to the special case of Problem 1.1 in which $f = \iota_E$, $(\forall i \in I)\ g_i = \phi_i \circ d_{C_i}$, and $(\forall j \in J)\ h_j = \psi_j \circ d_{C_j}$. To solve it with a fully proximal algorithm, one can use Lemma 2.1 for the nonsmooth terms, and Example 2.3 or Example 2.7 for the smooth ones.

Remark 4.9. The proposed variational formulation (4.10) is also of interest beyond the field of image recovery. For instance, the set-theoretic Fermat-Weber (a.k.a. Heron) problem arising in location theory is to [50]

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{i\in I} d_{C_i}(x). \tag{4.16}$$

Problem 4.7 therefore provides variants and generalizations of this formulation.

4.4.2. Application to image reconstruction from phase. This numerical example revolves around the classical problem of recovering an image $\overline{x}$ from the observation of its Fourier phase $\theta = \angle\widehat{\overline{x}}$ [45, 46]. The original 512×512 image $\overline{x}$ is shown in Fig. 4.8(a). The problem is modeled by Problem 4.6 as a convex feasibility problem with the following constraint sets.

- Mean pixel value: $C_1 = \big\{x \in \mathbb{R}^N \;\big|\; \langle x \mid 1\rangle = \mu\big\}$.
- Upper bound on the norm of the gradient: $Dx \in C_2$, with

  $$C_2 = \big\{y \in \mathbb{R}^N \times \mathbb{R}^N \;\big|\; \|y\|_2 \leqslant \eta\big\}, \tag{4.17}$$

  where D is as in (4.6).
- Phase: $C_3 = \big\{x \in \mathbb{R}^N \;\big|\; \angle\widehat{x} = \theta\big\}$.
- Proximity to the reference image r of Fig. 4.8(b): $C_4 = \big\{x \in \mathbb{R}^N \;\big|\; \|x - r\|_2 \leqslant \xi\big\}$. The image r is a blurred and noise corrupted version of $\overline{x}$, which is further degraded by saturation (the pixel values beyond 130 are clipped to 130) and the addition of a local high intensity noise on a rectangular area around the right eye. The only information available to the user is the bound ξ on the distance of r to the true image.
- Pixel range: $C_5 = [0, 255]^N$.

Now set $E = C_5$, $K = \{1, 2, 3, 4\}$, $\mathcal{H} = \mathcal{G}_1 = \mathcal{G}_3 = \mathcal{G}_4 = \mathbb{R}^N$, $\mathcal{G}_2 = \mathbb{R}^N \times \mathbb{R}^N$, and $L_2 = D$ in Problem 4.6. Then the feasibility problem (4.9) amounts to finding an image $x \in C_1 \cap D^{-1}(C_2) \cap C_3 \cap C_4 \cap C_5$. Because of inaccuracies in the values of μ, ξ, θ, and η, this problem turns out to be inconsistent and we therefore turn to the formulation of Problem 4.7. To ensure the robustness of the model to possible outliers, we adopt a constrained Huber framework, namely

$$\underset{x\in C_5}{\text{minimize}} \quad \mathfrak{h}_{\rho_1}(d_{C_1}(x)) + \mathfrak{h}_{\rho_2}(d_{C_2}(Dx)) + \mathfrak{h}_{\rho_3}(d_{C_3}(x)) + \mathfrak{h}_{\rho_4}(d_{C_4}(x)), \tag{4.18}$$

where the functions $(\mathfrak{h}_{\rho_i})_{1\leqslant i\leqslant 4}$ are defined in (2.19) and $\rho_1 = \rho_2 = \rho_3 = 1000$ and $\rho_4 = 5000$. Two versions of Problem 1.1 are employed.

Problem 4.10. In Problem 1.1, set $f = \iota_{C_5}$, $I = \varnothing$, $J = \{1, 2, 3, 4\}$, $h_1 = \mathfrak{h}_{\rho_1} \circ d_{C_1}$, $L_1 = \mathrm{Id}$, $h_2 = \mathfrak{h}_{\rho_2} \circ d_{C_2}$, $L_2 = D$, $h_3 = \mathfrak{h}_{\rho_3} \circ d_{C_3}$, $L_3 = \mathrm{Id}$, $h_4 = \mathfrak{h}_{\rho_4} \circ d_{C_4}$, and $L_4 = \mathrm{Id}$.

Problem 4.11 (fully proximal). In Problem 1.1, set $f = \iota_{C_5}$, $I = \{1, 2, 3, 4\}$, $J = \varnothing$, $g_1 = \mathfrak{h}_{\rho_1} \circ d_{C_1}$, $L_1 = \mathrm{Id}$, $g_2 = \mathfrak{h}_{\rho_2} \circ d_{C_2}$, $L_2 = D$, $g_3 = \mathfrak{h}_{\rho_3} \circ d_{C_3}$, $L_3 = \mathrm{Id}$, $g_4 = \mathfrak{h}_{\rho_4} \circ d_{C_4}$, and $L_4 = \mathrm{Id}$.

(a) (b) (c)

Figure 4.8. *(a) Original image $\overline{x}$. (b) Reference image r. (c) Reconstructed image (all algorithms yield visually equivalent images).*

We apply to Problems 4.10 and 4.11 Algorithms 3.7, 3.9, and 3.11 with all initial vectors set to 0. The following parameters are used, where $\beta = \sqrt{\sum_{i\in I}\|L_i\|^2 + \sum_{j\in J}\mu_j\|L_j\|^2}$ (these parameters were found to optimize the performance of each algorithm):

- Algorithm 3.7-S (with smooth terms for Problem 4.10): $\gamma_n \equiv 0.99/\beta$.
- Algorithm 3.7-P (fully proximal for Problem 4.11): $\gamma_n \equiv 0.99/\beta$.
- Algorithm 3.9-S (with smooth terms for Problem 4.10): $\tau_n \equiv 1.99/\beta$.
- Algorithm 3.9-P (fully proximal for Problem 4.11): $\sigma_{1,n} \equiv 1/(1.1\beta)$, $\sigma_{2,n} \equiv 1/(1.1\beta)$, $\sigma_{3,n} \equiv 1/(1.1\beta)$, $\sigma_{4,n} \equiv 1/(1.1\beta)$, and $\tau_n \equiv 1/\beta$.
- Algorithm 3.11-S (with smooth terms for Problem 4.10): $\gamma_n \equiv 0.5$, $\mu_{1,n} \equiv 0.99$, $\mu_{2,n} \equiv 0.99$, $\mu_{3,n} \equiv 0.99$, $\mu_{4,n} \equiv 0.99$, and $\lambda_n \equiv 1.9$.
- Algorithm 3.11-P (fully proximal for Problem 4.11): $\gamma_n \equiv 0.25$, $\mu_{1,n} \equiv 2.0$, $\mu_{2,n} \equiv 2.0$, $\mu_{3,n} \equiv 0.5$, $\mu_{4,n} \equiv 2.0$, and $\lambda_n \equiv 1.9$.

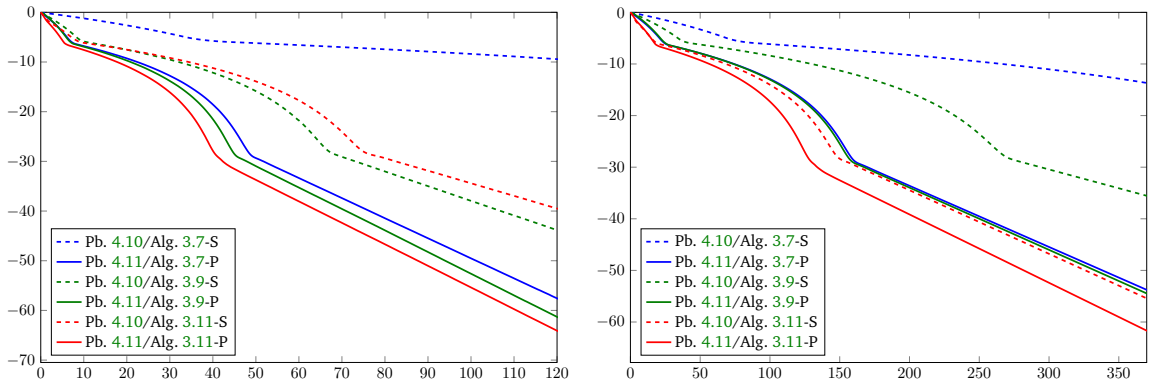

Figure 4.9. *Left: Normalized distance in dB to the asymptotic image produced by each algorithm versus execution time in seconds. Right: Normalized distance in dB to the asymptotic image produced by each algorithm versus iteration number.*

The gradient and proximity operators of the functions $(g_i)_{1\leqslant i\leqslant 4}$ are derived directly from

Example 2.3. They involve the projection operators $(\mathrm{proj}_{C_i})_{1\leqslant i\leqslant 4}$, which can be found in [23, 64], as well as the proximity operator of $\mathfrak{h}_\rho$, which can be found in [9, Example 24.9]. As seen in Fig. 4.8(c), despite the inconsistencies in the a priori knowledge, the reconstructed image captures important features of the original image. The results of Fig. 4.9 show the faster convergence of the fully proximal algorithms compared to the gradient-based ones.

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, 59 (2017), pp. 415–431.

[2] A. Alotaibi, P. L. Combettes, and N. Shahzad, *Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set,* SIAM J. Optim., 24 (2014), pp. 2076–2095.

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

[4] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, *Sparse/robust estimation and Kalman smoothing with nonsmooth log-concave densities: Modeling, computation, and theory,* J. Mach. Learn. Res., 14 (2013), pp. 2689–2728.

[5] 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., 9 (2016), pp. 842–868.

[6] J.-P. Aubin and A. Cellina, *Differential Inclusions: Set-Valued Maps and Viability Theory,* Springer-Verlag, New York, 1984.

[7] J.-F. Aujol and C. Dossal, *Stability of over-relaxations for the forward-backward algorithm, application to FISTA,* SIAM J. Optim., 25 (2015), pp. 2408–2433.

[8] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems,* SIAM Rev., 38 (1996), pp. 367–426.

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

[10] H. H. Bauschke, P. L. Combettes, and D. R. Luke, *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces,* J. Approx. Theory, 127 (2004), pp. 178–192.

[11] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems,* SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[12] R. Bellman, R. E. Kalaba, and J. A. Lockett, *Numerical Inversion of the Laplace Transform: Applications to Biology, Economics Engineering, and Physics,* Elsevier, New York, 1966.

[13] 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., 2 (2009), pp. 485–508.

[14] C. L. Byrne, *Iterative Optimization in Inverse Problems,* CRC Press, Boca Raton, FL, 2014.

[15] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld, *The multiple-sets split feasibility problem and its applications for inverse problems,* Inverse Problems, 21 (2005), pp. 2071–2084.

[16] Y. Censor and M. Zaknoon, *Algorithms and convergence results of projection methods for inconsistent feasibility problems: A review,* Pure Appl. Funct. Anal., 3 (2018), pp. 565–586.

[17] 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., 15 (2011), pp. 185–201.

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

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

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

[21] 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., 8 (2015), pp. 2662–2682.
[22] P. L. COMBETTES, *Inconsistent signal feasibility problems: Least-squares solutions in a product space,* IEEE Trans. Signal Process., 42 (1994), pp. 2955–2966.
[23] P. L. COMBETTES, *The convex feasibility problem in image recovery*, in Advances in Imaging and Electron Physics, P. Hawkes, ed., vol. 95, Academic Press, New York, 1996, pp. 155–270.
[24] P. L. COMBETTES, *Solving monotone inclusions via compositions of nonexpansive averaged operators,* Optimization, 53 (2004), pp. 475–504.
[25] P. L. COMBETTES, *Perspective functions: Properties, constructions, and examples,* Set-Valued Var. Anal., 26 (2018), pp. 247–264.
[26] P. L. COMBETTES AND P. BONDON, Hard-constrained inconsistent signal feasibility problems, IEEE Trans. Signal Process., 47 (1999), pp. 2460–2468.
[27] P. L. COMBETTES, L. CONDAT, J.-C. PESQUET, AND B. C. VŨ, *A forward-backward view of some primal-dual optimization methods in image recovery,* Proc. IEEE Int. Conf. Image Process., Paris, France, Oct. 27–30, 2014, pp. 4141–4145.
[28] P. L. COMBETTES, DINH DŨNG, AND B. C. VŨ, *Dualization of signal recovery problems,* Set-Valued Var. Anal., 18 (2010), pp. 373–404.
[29] P. L. COMBETTES AND J. ECKSTEIN, *Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions,* Math. Program., B168 (2018), pp. 645–672.
[30] P. L. COMBETTES AND J.-C. PESQUET, *A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,* IEEE J. Select. Topics Signal Process., 1 (2007), pp. 564–574.
[31] P. L. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing,* in: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, H. H. Bauschke *et al.*, eds., Springer, New York, 2011, pp. 185–212.
[32] 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., 20 (2012), pp. 307–330.
[33] P. L. COMBETTES, S. SALZO, AND S. VILLA, *Consistent learning by composite proximal thresholding,* Math. Program., B167 (2018), pp. 99–127.
[34] P. L. COMBETTES AND B. C. VŨ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality,* Optimization, 63 (2014), pp. 1289–1318.
[35] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting,* Multiscale Model. Simul., 4 (2005), pp. 1168–1200.
[36] L. CONDAT, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,* J. Optim. Theory Appl., 158 (2013), pp. 460–479.
[37] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,* Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.
[38] C. DE MOL AND M. DEFRISE, *A note on wavelet-based inversion algorithms,* Contemp. Math., 313 (2002), pp. 85–96.
[39] M. GOLDBURG AND R. J. MARKS II, *Signal synthesis in the presence of an inconsistent set of constraints,* IEEE Trans. Circuits Syst., 32 (1985), pp. 647–663.
[40] B. HE AND X. YUAN, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective,* SIAM J. Imaging Sci., 5 (2012), pp. 119–149.
[41] M. HINTERMÜLLER AND G. STADLER, *An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration,* SIAM J. Sci. Comput., 28 (2006), pp. 1–23.
[42] P. J. HUBER, *Robust Statistics,* 1st ed., Wiley, New York, 1981.
[43] P. R. JOHNSTONE AND J. ECKSTEIN, *Projective splitting with forward steps: Asynchronous and block-iterative operator splitting,* arxiv, 2018. https://arxiv.org/pdf/1803.07043.pdf
[44] S. L. KEELING, *Total variation based convex filters for medical imaging,* Appl. Math. Comput., 139 (2003), pp. 101–119.
[45] D. KERMISCH, *Image reconstruction from phase information only,* J. Opt. Soc. Amer., 60 (1970), pp. 11–17.
[46] A. LEVI AND H. STARK, *Signal reconstruction from phase by projection onto convex sets,* J. Opt. Soc. Amer., 73 (1983), pp. 810–822.
[47] Y. LIU, Z. LIANG, AND J. MA, *Total variation-Stokes strategy for sparse-view X-ray CT image reconstruction,* IEEE Trans. Med. Imaging, 33 (2014), pp. 749–763.

[48] C. Louchet and L. Moisan, *Posterior expectation of the total variation model: Properties and experiments,* SIAM J. Imaging Sci., 6 (2013), pp. 2640–2684.
[49] B. Martinet, *Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox,* C. R. Acad. Sci. Paris, A274 (1972), pp. 163–165.
[50] B. S. Mordukhovich, N. M. Nam, and J. Salinas, *Solving a generalized Heron problem by means of convex analysis,* Amer. Math. Monthly, 119 (2012), pp. 87–99.
[51] J. J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien,* C. R. Acad. Sci. Paris Sér. A, 255 (1962), pp. 2897–2899.
[52] D. O'Connor and L. Vandenberghe, *Primal-dual decomposition by operator splitting and applications to image deblurring,* SIAM J. Imaging Sci., 7 (2014), pp. 1724–1754.
[53] N. Papadakis, G. Peyré, and E. Oudet, *Optimal transport with proximal splitting,* SIAM J. Imaging Sci., 7 (2014), pp. 212–238.
[54] H. Raguet and L. Landrieu, *Preconditioning of a generalized forward-backward splitting and application to optimization on graphs,* SIAM J. Imaging Sci., 8 (2015), pp. 2706–2739.
[55] C. Schnörr, *Unique reconstruction of piecewise-smooth images by minimizing strictly convex nonquadratic functionals,* J. Math. Imaging Vision, 4 (1994), pp. 189–198.
[56] M. I. Sezan and H. Stark, *Image restoration by convex projections in the presence of noise,* Appl. Opt., 22 (1983), pp. 2781–2789.
[57] I. Steinwart, *Consistency of support vector machines and other regularized kernel classifiers,* IEEE Trans. Inform. Theory, 51 (2005), pp. 128–142.
[58] M. Tofighi, O. Yorulmaz, K. Köse, D. C. Yildirim, R. Çetin-Atalay, and A. E. Çetin, *Phase and TV based convex sets for blind deconvolution of microscopic images,* IEEE J. Selected Topics Signal Process., 10 (2016), pp. 81–91.
[59] V. N. Vapnik, *The Nature of Statistical Learning Theory,* 2nd ed., Springer, New York, 2000.
[60] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators,* Adv. Comput. Math., 38 (2013), pp. 667–681.
[61] B. C. Vũ, *A variable metric extension of the forward-backward-forward algorithm for monotone operators,* Numer. Funct. Anal. Optim., 34 (2013), pp. 1–16.
[62] D. C. Youla, *Generalized image restoration by the method of alternating orthogonal projections,* IEEE Trans. Circuits Syst., 25 (1978), pp. 694–702.
[63] 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., 33 (1986), pp. 465–468.
[64] D. C. Youla and H. Webb, *Image restoration by the method of convex projections: Part 1 – theory,* IEEE Trans. Med. Imaging, 1 (1982), pp. 81–94.
[65] T. Zhang, *Statistical behavior and consistency of classification methods based on convex risk minimization,* Ann. Statist., 32 (2004), pp. 56–85.