A Douglas-Rachford Splitting Approach to Nonsmooth Convex Variational Signal Recovery

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

***Abstract*— Under consideration is the large body of signal recovery problems that can be formulated as the problem of minimizing the sum of two (not necessarily smooth) lower semicontinuous convex functions in a real Hilbert space. This generic problem is analyzed and a decomposition method is proposed to solve it. The convergence of the method, which is based on the Douglas-Rachford algorithm for monotone operator-splitting, is obtained under general conditions. Applications to non-Gaussian image denoising in a tight frame are also demonstrated.**

***Index Terms*— Convex optimization, denoising, Douglas-Rachford, frame, nondifferentiable optimization, Poisson noise, proximal algorithm, wavelets.**

I. INTRODUCTION

The objective of a signal recovery problem is to infer the original form of a signal $\overline{x}$ from the observation of signals which are physically related to it via some known mathematical models. This broad class of inverse problems encompasses problems such as signal reconstruction, signal restoration, and signal denoising [2], [17], [46], [50]. In the early days, it was customary to divide recovery problems into linear and nonlinear problems, the former being naturally easier to analyze and solve. For instance, Youla showed in [53] that a wide range of classical signal recovery problems could be reduced to the problem of finding a common point of two affine subspaces of a Hilbert space and solved via alternating affine projection methods (see [15] for further historical background). Likewise, in the area of computed tomography, most of the algorithms developed in the 1970s revolved around iterative affine projection methods [37].

In mathematics, the methodological barrier between linear and nonlinear analysis started to disappear in the 1950s and 1960s with the development of various well-structured nonlinear theories (e.g., fixed point theory, game theory, monotone operator theory, or convex analysis) [4], [58]. In signal recovery, the emergence of convex analysis as a powerful nonlinear framework crystallized around the paper [55] in which Youla and Webb, bringing to light work from the Russian school from the 1960s [15], replaced the affine projection operators of the alternating projection method with convex projection operators to obtain the so-called POCS (Projection Onto Convex Sets) algorithm. POCS has been used in scores of signal processing problems [15], [17], [47] and has been extended to more flexible projection algorithms [16], [18] that have found further applications, e.g., [1], [12], [34], [35], [43], [52], [56]. Recall that, if C is a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$ with norm $\|\cdot\|$, then, for every $x \in \mathcal{H}$, there exists a unique point $P_C x$ in $\mathcal{H}$ – called the *projection* of x onto C – which satisfies the best approximation property

$$P_C x \in C \quad \text{and} \quad (\forall y \in C) \quad \|x - P_C x\| \leq \|x - y\|. \tag{1}$$

Alternatively, $P_C x$ is the unique solution to the variational problem

$$\min_{y \in \mathcal{H}} \iota_C(y) + \frac{1}{2}\|x - y\|^2, \tag{2}$$

where ι_C is the *indicator function* of C, i.e.,

$$(\forall y \in \mathcal{H}) \quad \iota_C(y) = \begin{cases} 0, & \text{if } y \in C; \\ +\infty, & \text{otherwise.} \end{cases} \tag{3}$$

Convex projection methods exploit the remarkable properties of projection operators [6], [19] and, in order to broaden the scope of these methods, it is natural to introduce more general operators with similar properties. Such an extension was proposed by Moreau in 1962 [39]. Under the above assumptions, ι_C belongs to the class $\Gamma_0(\mathcal{H})$ of all functions from $\mathcal{H}$ to $]-\infty, +\infty]$ which are lower semicontinuous, convex, and not identically equal to $+\infty$. Moreau observed that, for every $x \in \mathcal{H}$, the variational problem obtained by replacing ι_C with an arbitrary function $f \in \Gamma_0(\mathcal{H})$ in (2), namely,

$$\min_{y \in \mathcal{H}} f(y) + \frac{1}{2}\|x - y\|^2, \tag{4}$$

admits a unique solution, which will be denoted by $\operatorname{prox}_f x$. The so-called *proximity operator* $\operatorname{prox}_f \colon \mathcal{H} \to \mathcal{H}$ thus defined generalizes the notion of a convex projection operator in the sense that $P_C = \operatorname{prox}_{\iota_C}$, and it moreover possesses most of its attractive properties [40]. As a result, many projection methods (in particular POCS) can be extended to proximal methods by replacing projection operators with proximity operators [7], [21].

In signal recovery, the use of proximity operators seems to originate in [20]. Their main advantage is to model various operations on signals beyond convex projections, e.g., nonexpansive self-adjoint linear transformations or soft thresholding operations (see [24] and [25] for details and additional examples). They also naturally arise in the analysis and the numerical solution of signal recovery problems. In this regard, it is shown

Manuscript received January 30, 2007; revised June 30, 2007; accepted September 3, 2007. The guest editor coordinating the review of this manuscript and approving it for publication was Dr Timothy Davidson.

P. L. Combettes is with the Laboratoire Jacques-Louis Lions – UMR CNRS 7598, Faculté de Mathématiques, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France. E-mail: plc@math.jussieu.fr.

J.-C. Pesquet is with the Institut Gaspard Monge and UMR–CNRS 8049, Université de Paris-Est Marne-la-Vallée, 77454 Marne la Vallée Cedex 2, France. E-mail: pesquet@univ-mlv.fr.

in [25] (see also [13] and [24] for further developments) that a number of apparently unrelated recovery formulations (e.g., constrained least-squares problems, multiresolution sparse regularization problems, Fourier regularization problems, geometry/texture image decomposition problems, hard-constrained inconsistent feasibility problems, split feasibility problems, as well as certain maximum *a posteriori* problems) fit the following simple variational format.

Problem 1 Let f_1 and f_2 be two functions in $\Gamma_0(\mathcal{H})$ such that f_2 is differentiable on $\mathcal{H}$ with a β-Lipschitz continuous gradient for some $\beta \in \,]0,+\infty[$. The objective is to minimize $f_1 + f_2$ over $\mathcal{H}$.

Investigating this generic formulation makes it possible to derive existence, uniqueness, and characterization results in a unified and standardized fashion.

Proposition 2 [25, Proposition 3.1]

i) *Problem 1 possesses at least one solution if* $\lim_{\|x\|\to+\infty} f_1(x) + f_2(x) = +\infty$.
ii) *Problem 1 possesses at most one solution if* $f_1 + f_2$ *is strictly convex, as is the case when* f_1 *or* f_2 *is strictly convex.*
iii) *Let* $x \in \mathcal{H}$ *and let* $\gamma \in \,]0,+\infty[$. *Then the following statements are equivalent.*
   a) x *solves Problem 1.*
   b) $x = \operatorname{prox}_{\gamma f_1}(x - \gamma\nabla f_2(x))$.

The fixed point characterization in Proposition 2iii) suggests a numerical method for solving Problem 1. This observation led to the following convergence result, in which the sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ stand for some tolerance in the inexact evaluations of the proximity operator of f_1 and of the gradient of f_2.

Theorem 3 [25, Theorem 3.4(i)] *Suppose that Problem 1 possesses at least one solution. Let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,+\infty[$ *such that* $\inf_{n\in\mathbb{N}}\gamma_n > 0$ *and* $\sup_{n\in\mathbb{N}}\gamma_n < 2/\beta$, *let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,1]$ *such that* $\inf_{n\in\mathbb{N}}\lambda_n > 0$, *and let* $(a_n)_{n\in\mathbb{N}}$ *and* $(b_n)_{n\in\mathbb{N}}$ *be sequences in* $\mathcal{H}$ *such that* $\sum_{n\in\mathbb{N}}\|a_n\| < +\infty$ *and* $\sum_{n\in\mathbb{N}}\|b_n\| < +\infty$. *Fix* $x_0 \in \mathcal{H}$ *and, for every* $n \in \mathbb{N}$, *set*

$$x_{n+1} = x_n + \lambda_n\Big(\operatorname{prox}_{\gamma_n f_1}\big(x_n - \gamma_n(\nabla f_2(x_n) + b_n)\big) + a_n - x_n\Big). \quad (5)$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 1.*

As discussed in [25], Theorem 3 extends and provides a simplified convergence analysis for a variety of existing iterative methods, such as the projected Landweber method [32], the alternating projection method [54], the parallel projection method [16], and the iterative soft thresholding method [27]. In [13], it also served as a basis to solve certain inverse problems posed on frames which were modeled as special instances of Problem 1 in the space of frame coefficients.

Despite its relatively broad range of applications, Problem 1 fails to cover the important situations in which f_2 is differentiable with a non-Lipschitz gradient, or not differentiable at all, or simply not finite everywhere. Nondifferentiability arises for instance in the problem of minimizing the total variation of a signal over a convex set, in the problem of minimizing the sum of two set-distance functions, in problems involving maxima of convex functions, or in Tykhonov-like problems with L^1 norms. The objective of the present paper is to relax the assumptions made on f_2 in Problem 1. The nonsmooth problem under consideration is formulated as follows (see Section II-B for the notation "cone" and "dom").

Problem 4 Let f_1 and f_2 be two functions in $\Gamma_0(\mathcal{H})$ such that

$$\operatorname{cone}(\operatorname{dom} f_1 - \operatorname{dom} f_2) \text{ is a closed vector subspace of } \mathcal{H}. \quad (6)$$

The objective is to minimize $f_1 + f_2$ over $\mathcal{H}$.

As will be seen in Proposition 14, the so-called qualification condition (6) is quite mild and it makes Problem 1 a special case of Problem 4.

Neither the differentiability of f_1 nor of f_2 is required in Problem 4. In this context, the algorithm described in (5) cannot be used and an alternative must be found. In finite-dimensional spaces, if both functions are finite, the standard subgradient method of [45, Chapter 2] can be used, but it is known to be slow due to its vanishing step-sizes. Another possibility is the adaptive level set method of [22], but it is tailored to problems in which one of the two functions is an indicator function. Our main contribution will be to propose an algorithm based on the Douglas-Rachford splitting method for monotone operators [21], [31], [38]. The chief advantage of a splitting method is to activate the functions f_1 and f_2 separately. Thus, the proposed iteration will use the operators $\operatorname{prox}_{\gamma f_1}$ and $\operatorname{prox}_{\gamma f_2}$, which are typically much easier to implement than the operator $\operatorname{prox}_{\gamma(f_1+f_2)}$ as in the standard proximal point algorithm for nonsmooth problems [21]

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma(f_1+f_2)}x_n. \quad (7)$$

The remainder of the paper is organized as follows. In Section II, we define our notation and provide the necessary mathematical background. Problem 4 is then analyzed in Section III and the algorithm proposed to solve it is introduced in Section IV, where a convergence result is established. Applications to non-Gaussian image denoising using a tight frame representation are addressed in Section V. Numerical results reporting on Laplace additive noise and Poisson data experiments are provided in Section VI. The appendices contain the proofs of certain technical results.

II. Notation and theoretical tools

In this section, we provide the necessary background on monotone operators, convex analysis, and proximity operators. Throughout this paper, $\mathcal{H}$ is a real Hilbert space with identity operator Id, scalar product $\langle\cdot\mid\cdot\rangle$, norm $\|\cdot\|$, and distance d.

A. *Monotone operators* [5]

Let $2^{\mathcal{H}}$ denote the class of all subsets of $\mathcal{H}$ and let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The set of *zeros* of A is

$$\operatorname{zer} A = \big\{x \in \mathcal{H} \;\big|\; 0 \in Ax\big\}. \tag{8}$$

We say that A is *monotone* if, for every (x,u) and (y,v) in $\mathcal{H}\times\mathcal{H}$ such that $u \in Ax$ and $v \in Ay$,

$$\langle x-y \mid u-v\rangle \geq 0. \tag{9}$$

Suppose that A is monotone. Then A is declared *maximal monotone* when the following property is satisfied for every $(x,u) \in \mathcal{H}\times\mathcal{H}$: if (9) holds for every $(y,v) \in \mathcal{H}\times\mathcal{H}$ such that $v \in Ay$, then $u \in Ax$. Now suppose that A is maximal monotone. Then, for every $x \in \mathcal{H}$, there exists a unique point $J_A x \in \mathcal{H}$ such that

$$x - J_A x \in A(J_A x). \tag{10}$$

The single-valued operator $J_A\colon \mathcal{H} \to \mathcal{H}$ thus defined is called the *resolvent* of A.

B. *Convex analysis* [57]

The *distance* function of a nonempty subset C of $\mathcal{H}$ is $d_C\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \inf_{y\in C}\|x-y\|$, its *indicator function* is defined in (3), its *support function* is $\sigma_C\colon \mathcal{H} \to \left]-\infty,+\infty\right] : u \mapsto \sup_{x\in C}\langle x \mid u\rangle$, and its *conical hull* is

$$\operatorname{cone} C = \bigcup_{\lambda>0}\big\{\lambda x \;\big|\; x \in C\big\}. \tag{11}$$

The *interior* of a set $C \subset \mathcal{H}$ is denoted by $\operatorname{int} C$ and its *relative interior* by $\operatorname{rint} C$ [42, Section 6]. The *domain* of a function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is $\operatorname{dom} f = \big\{x \in \mathcal{H} \;\big|\; f(x) < +\infty\big\}$. $\Gamma_0(\mathcal{H})$ is the class of all lower semicontinuous convex functions from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$ with nonempty domains. The *subdifferential* of $f \in \Gamma_0(\mathcal{H})$ at $x \in \mathcal{H}$ is the set

$$\partial f(x) = \big\{u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H})\;\langle y-x \mid u\rangle + f(x) \leq f(y)\big\}. \tag{12}$$

If $f \in \Gamma_0(\mathcal{H})$ is Gâteaux differentiable at $x \in \mathcal{H}$ with gradient $\nabla f(x)$, then $\partial f(x) = \{\nabla f(x)\}$. A tutorial account of subdifferentials can be found in [18].

Lemma 5 [57, Theorem 3.1.11] *Let $f \in \Gamma_0(\mathcal{H})$. Then ∂f is maximal monotone.*

We shall also require the following trivial, yet fundamental, consequence of (12), which generalizes the well-known fact that critical points of differentiable convex functions are minimizers.

Lemma 6 *Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \mathcal{H}$. Then x is a minimizer of f if and only if $x \in \operatorname{zer}\partial f$.*

C. *Proximity operators*

The foundational paper of the theory of proximity operators is [40]. This paper contains also many useful properties of proximity operators. Further properties and concrete examples are given in [13], [24], [25].

Recall that, for every $f \in \Gamma_0(\mathcal{H})$ and every $x \in \mathcal{H}$, $\operatorname{prox}_f x$ is the *unique* solution to the minimization problem (4). We first provide a subdifferential characterization of proximity operators.

Lemma 7 [40, Proposition 6.a] *Let $f \in \Gamma_0(\mathcal{H})$, let $x \in \mathcal{H}$, and let $p \in \mathcal{H}$. Then $p = \operatorname{prox}_f x \Leftrightarrow x - p \in \partial f(p)$.*

The following fact results from Lemma 5, (10), and Lemma 7. It states that prox_f is the resolvent of a maximal monotone operator, namely of the subdifferential of f.

Lemma 8 *Let $f \in \Gamma_0(\mathcal{H})$. Then $\operatorname{prox}_f = J_{\partial f}$.*

The next property is essential to establish the convergence of various algorithms based on proximity operators [21].

Lemma 9 [40, Proposition 5.b] *Let $f \in \Gamma_0(\mathcal{H})$. Then prox_f is nonexpansive: $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\;\|\operatorname{prox}_f x - \operatorname{prox}_f y\| \leq \|x-y\|$.*

In signal recovery problems, information is often available to penalize individually the coefficients of the original signal in an orthonormal basis [13], [24], [25], [27]. The following decomposition formula makes the implementation of the resulting proximity operator particularly convenient in such instances.

Lemma 10 [13, Remark 3.2(ii) and Proposition 2.10] *Set*

$$\Phi\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \sum_{k\in\mathbb{K}} \varphi_k(\langle x \mid o_k\rangle), \tag{13}$$

where:

i) $\varnothing \neq \mathbb{K} \subset \mathbb{N}$*;*
ii) $(o_k)_{k\in\mathbb{K}}$ *is an orthonormal basis of* $\mathcal{H}$*;*
iii) $(\varphi_k)_{k\in\mathbb{K}}$ *are functions in* $\Gamma_0(\mathbb{R})$*;*
iv) *Either* $\mathbb{K}$ *is finite, or there exists a subset* $\mathbb{L}$ *of* $\mathbb{K}$ *such that:*
   a) $\mathbb{K} \smallsetminus \mathbb{L}$ *is finite;*
   b) $(\forall k \in \mathbb{L})\;\varphi_k \geq \varphi_k(0) = 0$.

Then $\Phi \in \Gamma_0(\mathcal{H})$ and

$$(\forall x \in \mathcal{H})\quad \operatorname{prox}_\Phi x = \sum_{k\in\mathbb{K}} \big(\operatorname{prox}_{\varphi_k}\langle x \mid o_k\rangle\big) o_k. \tag{14}$$

We shall also exploit the following properties, which appear to be new.

Proposition 11 *Let $\mathcal{G}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{G})$, and let $L\colon \mathcal{H} \to \mathcal{G}$ be a bounded linear operator with closed range $\operatorname{ran} L$. Suppose that the composition of L and L^* satisfies $L \circ L^* = \nu\mathrm{Id}$, for some $\nu \in \left]0,+\infty\right[$, and that*

$$\operatorname{cone}(\operatorname{dom} f - \operatorname{ran} L) \text{ is a closed vector subspace of } \mathcal{G}. \tag{15}$$

Then $f \circ L \in \Gamma_0(\mathcal{H})$ *and*

$$\operatorname{prox}_{f \circ L} = \operatorname{Id} + \nu^{-1} L^* \circ (\operatorname{prox}_{\nu f} - \operatorname{Id}) \circ L. \tag{16}$$

Proof: See Appendix I. ■

Proposition 12 *Let* $f \in \Gamma_0(\mathcal{H})$ *and let* C *be a closed convex subset of* $\mathcal{H}$ *such that* $C \cap \operatorname{dom} f \neq \varnothing$. *Then the following hold.*

i) $(\forall x \in \mathcal{H})\ \operatorname{prox}_f x \in C \Rightarrow \operatorname{prox}_{\iota_C + f} x = \operatorname{prox}_f x$.
ii) *Suppose that* $\mathcal{H} = \mathbb{R}$. *Then* $\operatorname{prox}_{\iota_C + f} = P_C \circ \operatorname{prox}_f$.

Proof: See Appendix II. ■

The following example shows that the conclusion of Proposition 12ii) fails when $\mathcal{H} \neq \mathbb{R}$.

Example 13 Let $\Lambda \colon \mathcal{H} \to \mathcal{H}$ be a positive, self-adjoint, bounded linear operator, let $u \in \mathcal{H}$ be such that $\|u\| = 1$, and let $\delta \in \mathbb{R}$. Set $f \colon y \mapsto \langle \Lambda y \mid y \rangle /2$, $C = \big\{ y \in \mathcal{H} \;\big|\; \langle y \mid u \rangle = \delta \big\}$, and $Q = (\operatorname{Id} + \Lambda)^{-1}$. Then, for every $x \in \mathcal{H}$,

$$\operatorname{prox}_{f + \iota_C} x = Qx + \frac{\delta - \langle Qx \mid u \rangle}{\langle Qu \mid u \rangle} Qu, \tag{17}$$

whereas $P_C(\operatorname{prox}_f x) = Qx + (\delta - \langle Qx \mid u \rangle)u$.

III. Analysis of Problem 4

We start our discussion by observing that the qualification condition (6) imposed in Problem 4 is a mild restriction that is satisfied in many standard scenarios.

Proposition 14 *Condition* (6) *is satisfied in each of the following cases.*

i) $\operatorname{dom} f_1 \cap \operatorname{int} \operatorname{dom} f_2 \neq \varnothing$ *or* $\operatorname{dom} f_2 \cap \operatorname{int} \operatorname{dom} f_1 \neq \varnothing$.
ii) f_1 *or* f_2 *is finite.*
iii) $\mathcal{H}$ *is finite-dimensional and the relative interiors of* $\operatorname{dom} f_1$ *and* $\operatorname{dom} f_2$ *have a nonempty intersection.*

Proof: See Appendix III. ■

It follows in particular from Proposition 14ii) that Problem 4 subsumes Problem 1. Thus, the examples of signal recovery problems discussed in [25] are covered by Problem 4. However, the following scenarios are covered by Problem 4, but not by Problem 1.

Example 15 $\mathcal{H}$ is either the Euclidean space $\mathbb{R}^K$ or the Sobolev space $\mathrm{H}^1(\Omega)$, where Ω is a bounded open domain of $\mathbb{R}^m$, f_1 is the L^1 norm or the indicator function of a nonempty closed convex set, and f_2 is the total variation. Then it follows from [23, Proposition 1] that the assumptions of Problem 4 are satisfied.

Example 16 C_1 and C_2 are nonempty closed convex sets, $\alpha > 0$, $1 \leq p < +\infty$, $f_1 = \alpha d_{C_1}^p$, and $f_2 = d_{C_2}$. Problem 4 then extends the standard convex feasibility problem, which corresponds to the case when $C_1 \cap C_2 \neq \varnothing$, i.e., to the problem of finding a signal that lies in both C_1 and C_2 [55].

Example 17 $\mathcal{H} = \mathrm{L}^2(\Omega)$, where Ω is a bounded open domain of $\mathbb{R}^m$, and the observed data assume the form $z = L\overline{x} + w$, where L is a bounded linear operator from $\mathcal{H}$ to a Hilbert space $\mathcal{G}$ and $w \in \mathcal{G}$ is an additive noise vector. Moreover, $f_1 \colon x \mapsto \|Lx - z\|_{\mathrm{L}^1}$ and $f_2 = \alpha \|\cdot\|_{\mathrm{L}^1}$, with $\alpha > 0$.

We now address the existence, uniqueness, and characterization of the solutions to Problem 4.

Proposition 18

i) *Problem 4 possesses at least one solution if* $\lim_{\|x\| \to +\infty} f_1(x) + f_2(x) = +\infty$.
ii) *Problem 4 possesses at most one solution if* $f_1 + f_2$ *is strictly convex, as is the case when* f_1 *or* f_2 *is strictly convex.*
iii) *Let* $x \in \mathcal{H}$ *and* $\gamma \in \left]0, +\infty\right[$, *and set, for every* $f \in \Gamma_0(\mathcal{H})$, $\operatorname{rprox}_f = 2\operatorname{prox}_f - \operatorname{Id}$. *Then the following statements are equivalent.*
   a) x *solves Problem 4.*
   b) $x \in \operatorname{zer}(\partial f_1 + \partial f_2)$.
   c) $x = \operatorname{prox}_{\gamma f_2} y$, *where* $y = \operatorname{rprox}_{\gamma f_1}(\operatorname{rprox}_{\gamma f_2} y)$.

Proof: See Appendix IV. ■

IV. Douglas-Rachford splitting algorithm

This algorithm was originally proposed in [30] for solving matrix equations of the form $u = Ax + Bx$, where A and B are positive-definite matrices, in connection with the numerical solution of partial differential equations [51]. The method was extended to the nonlinear problem of finding a zero of the sum of two maximal monotone operators in Hilbert spaces in [38] and revisited in [31]; further properties are discussed in [8]. To the best of our knowledge, the most general result on the convergence of the Douglas-Rachford algorithm is the following.

Theorem 19 [21, Corollary 5.2] *Let* $\gamma \in \left]0, +\infty\right[$, *let* $(\lambda_n)_{n \in \mathbb{N}}$ *be a sequence in* $\left]0, 2\right[$, *and let* $(a_n)_{n \in \mathbb{N}}$ *and* $(b_n)_{n \in \mathbb{N}}$ *be sequences in* $\mathcal{H}$. *Suppose that the following hold.*

i) $\operatorname{zer}(A + B) \neq \varnothing$.
ii) $\sum_{n \in \mathbb{N}} \lambda_n (2 - \lambda_n) = +\infty$.
iii) $\sum_{n \in \mathbb{N}} \lambda_n (\|a_n\| + \|b_n\|) < +\infty$.

Take $x_0 \in \mathcal{H}$ *and set, for every* $n \in \mathbb{N}$,

$$\begin{cases} x_{n+\frac{1}{2}} = J_{\gamma B} x_n + b_n \\ x_{n+1} = x_n + \lambda_n \Big(J_{\gamma A}\big(2x_{n+\frac{1}{2}} - x_n\big) + a_n - x_{n+\frac{1}{2}} \Big). \end{cases} \tag{18}$$

Then $(x_n)_{n \in \mathbb{N}}$ *converges weakly to some point* $x \in \mathcal{H}$ *and* $J_{\gamma B} x \in \operatorname{zer}(A + B)$.

An important feature of Algorithm (18) is that it proceeds by *splitting* in the sense that the operators A and B are used in separate steps: in the first step, only the resolvent of B is required to obtain $x_{n+\frac{1}{2}}$ and, in the second step, only resolvent of A is required to obtain x_{n+1}. Now set $A = \partial f_1$ and $B = \partial f_2$. Then Lemma 5 asserts that A and B are maximal monotone. Moreover, in view of Proposition 18iii), Problem 4 is equivalent to finding a point in $\operatorname{zer}(A + B)$.

Hence, using Lemma 8, we derive at once from Theorem 19 the following convergence result on a proximal method for solving Problem 4.

Theorem 20 *Let* $\gamma \in \left]0,+\infty\right[$, *let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $\left]0,2\right[$, *and let* $(a_n)_{n\in\mathbb{N}}$ *and* $(b_n)_{n\in\mathbb{N}}$ *be sequences in* $\mathcal{H}$. *Suppose that the following hold.*

i) *Problem 4 admits at least one solution.*
ii) $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$.
iii) $\sum_{n\in\mathbb{N}} \lambda_n(\|a_n\| + \|b_n\|) < +\infty$.

Take $x_0 \in \mathcal{H}$ *and set, for every* $n \in \mathbb{N}$,

$$\begin{cases} x_{n+\frac{1}{2}} = \operatorname{prox}_{\gamma f_2} x_n + b_n \\ x_{n+1} = x_n + \lambda_n\Big(\operatorname{prox}_{\gamma f_1}\big(2x_{n+\frac{1}{2}} - x_n\big) + a_n - x_{n+\frac{1}{2}}\Big). \end{cases} \tag{19}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to some point* $x \in \mathcal{H}$ *and* $\operatorname{prox}_{\gamma f_2} x$ *is a solution to Problem 4.*

Algorithm (19) naturally inherits the splitting property of the Douglas-Rachford iteration (18), which makes it particularly attractive in comparison with standard proximal methods such as (7). Thus, each iteration n is decomposed into two main steps: the current iterate is x_n and the function f_2 is first utilized to compute $x_{n+\frac{1}{2}}$; the function f_1 is then utilized to produce the update x_{n+1}. Note that Algorithm (19) allows for the inexact implementation of these two proximal steps via the incorporation of the error terms a_n and b_n. Moreover, a variable relaxation parameter λ_n provides added flexibility (it can of course be set to a constant value in $\left]0,2\right[$). It is important to note that the solution to Problem 4 is obtained as the image under $\operatorname{prox}_{\gamma f_2}$ of the weak limit of $(x_n)_{n\in\mathbb{N}}$ and that, in general, little is known about the asymptotic behavior of $\big(\operatorname{prox}_{\gamma f_2} x_n\big)_{n\in\mathbb{N}}$ unless $\operatorname{prox}_{\gamma f_2}$ is weakly continuous. This property is notably satisfied when $\mathcal{H}$ is finite-dimensional since weak and strong convergence coincide in this case and $\operatorname{prox}_{\gamma f_2}$ is continuous by Lemma 9. Thus, we derive from Theorem 20 that $\big(\operatorname{prox}_{\gamma f_2} x_n\big)_{n\in\mathbb{N}}$ converges to a solution and, since $b_n \to 0$ in (19), we obtain the following result, which is immediately relevant to digital signal processing.

Corollary 21 *Suppose that* $\mathcal{H}$ *is finite-dimensional. Then the sequence* $(x_{n+\frac{1}{2}})_{n\in\mathbb{N}}$ *in Theorem 20 converges to a solution to Problem 4.*

V. Applications to non-Gaussian denoising in a tight frame

As discussed in Section III, the proposed approach is applicable to a wide array of signal recovery problems. In this section, the focus is placed on image denoising in the presence of non-Gaussian noise. Let us emphasize that variational problems posed on frames were already considered in [13]. However, the problems investigated in [13] were covered by Problem 1, whereas the problems presented below are more general instances of Problem 4 that contain no smooth term.

A. Problem statement

We consider the problem of recovering an M-dimensional signal $\overline{y}$ in the Euclidean space $\mathcal{G} = \mathbb{R}^M$ from a noisy observation $z \in \mathcal{G}$. No specific degradation model is assumed and, in particular, the noise perturbation need not be additive. It is assumed that *a priori* information is available, that constrains $\overline{y}$ to lie in a closed convex subset C of $\mathcal{G}$.

In previous work on signal denoising (see [3], [9], [24], [44] and the references therein), the representation of $\overline{y}$ in an appropriate orthonormal basis has been shown to be helpful in the modeling of various properties of the original signal $\overline{y}$, e.g., regularity or sparsity. We place ourselves in a more general framework by considering redundant representations built from overcomplete sets (i.e., frames) of vectors in $\mathcal{G}$. More precisely, we assume that information is available on the coefficients $\overline{x} = (\overline{\xi}_k)_{1\le k\le K}$ $(K \geq M)$ of $\overline{y}$ in a tight frame $(e_k)_{1\le k\le K}$ of $\mathcal{G}$. Recall that, if we denote by $\mathcal{H}$ the Euclidean space $\mathbb{R}^K$, the frame operator associated with $(e_k)_{1\le k\le K}$ is the injective linear operator

$$F\colon \mathcal{G} \to \mathcal{H}\colon y \mapsto (\langle y \mid e_k\rangle)_{1\le k\le K}, \tag{20}$$

the adjoint of which is the surjective linear operator

$$F^*\colon \mathcal{H} \to \mathcal{G}\colon (\xi_k)_{k\in\mathbb{K}} \mapsto \sum_{k=1}^{K} \xi_k e_k. \tag{21}$$

For a tight frame, we have

$$F^* \circ F = \nu \mathrm{Id}, \quad \text{where} \quad \nu \in \left]0,+\infty\right[. \tag{22}$$

A simple example of a tight frame is the union of q orthonormal bases, in which case $\nu = q$. Discretized versions of curvelets [11], dual-tree wavelet decompositions [14], Gabor frames [26], [49], and contourlets [29] constitute tight frames under suitable hypotheses. For a detailed account of frame theory, see [36].

Our strategy will be to produce an estimate $\tilde{x}$ of the frame coefficient vector $\overline{x}$ in $\mathcal{H} = \mathbb{R}^K$ through the maximum *a posteriori* approach described below, which will turn out to be a special case of Problem 4. We shall then construct $\tilde{x}$ via Algorithm (19) and Corollary 21 and, in turn, obtain an estimate of $\overline{y}$ as $\tilde{y} = F^*\tilde{x}$. Let us emphasize that the resulting variational formulations will, in general, be outside of the scope of Problem 1 due to the potential lack of differentiability of both functions (see Section VI for examples).

B. Bayesian formulation

In the remainder of Section V, the following assumptions are made on the constituents of the problem.

Assumption 22

i) The vectors $\overline{x}$, $\overline{y}$, and $z = (\zeta_m)_{1\le m\le M}$ are realizations of real-valued random vectors $\overline{X}$, $\overline{Y} = (\overline{\Upsilon}_m)_{1\le m\le M}$, and $Z = (Z_m)_{1\le m\le M}$, respectively (all random variables are defined on the same probability space).
ii) Given $\overline{Y} = (\eta_m)_{1\le m\le M}$, the components $(Z_m)_{1\le m\le M}$ of Z are conditionally independent random variables

which are either discrete, with conditional probability mass functions $\big(\mu_{Z_m|\overline{\Upsilon}_m=\eta_m}\big)_{1\le m\le M}$, or absolutely continuous, in which case we also denote by $\big(\mu_{Z_m|\overline{\Upsilon}_m=\eta_m}\big)_{1\le m\le M}$ their probability density functions.

iii) For every $m\in\{1,\ldots,M\}$, we have $\mu_{Z_m|\overline{\Upsilon}_m=\cdot}(\zeta_m)\propto\exp(-\psi_m(\cdot))$, where $\psi_m\in\Gamma_0(\mathbb{R})$.

iv) The components of $\overline{X}$ are independent with densities $(\exp(-\phi_k(\cdot))/\int_{-\infty}^{+\infty}\exp(-\phi_k(\eta))d\eta)_{1\le k\le K}$, where $(\phi_k)_{1\le k\le K}$ are finite convex functions on the real line.

v) The Lebesgue measure of $(\operatorname{dom}\psi_1\times\cdots\times\operatorname{dom}\psi_M)\cap C$ is nonzero.

Note that Assumption 22iii) is satisfied in particular in the case of an additive noise $Z_m-\overline{\Upsilon}_m$ which is independent of $\overline{\Upsilon}_m$ and possesses a upper-semicontinuous log-concave density, for every $m\in\{1,\ldots,M\}$. Various examples of such laws are provided in [13].

Under Assumption 22, a standard Bayesian approach for estimating $\overline{x}$ from z consists of applying a maximum *a posteriori* rule [10], [33], [48], which amounts to maximizing the posterior probability density $\mu_{\overline{X}|Z=z,\overline{Y}\in C}$ of $\overline{X}$. Using Bayes' formula, this is equivalent to solving the variational problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\quad -\ln\mu_{\overline{X}|\overline{Y}\in C}(x)-\ln\mu_{Z|\overline{Y}=F^*x}(z),\tag{23}$$

where $\mu_{Z|\overline{Y}=F^*x}$ is the conditional probability (mass or density) function of Z given $\overline{Y}=F^*x$. Note that, upon setting $x=(\xi_k)_{1\le k\le K}$ and using Assumption 22iv), we obtain

$$\mu_{\overline{X}|\overline{Y}\in C}(x)=\begin{cases}\vartheta^{-1}\prod_{k=1}^K\exp(-\phi_k(\xi_k)), & \text{if } F^*x\in C;\\ 0, & \text{otherwise,}\end{cases}\tag{24}$$

where ϑ is the probability that $\overline{Y}\in C$, which is nonzero by Assumption 22v). In addition, according to Assumptions 22ii) and 22iii),

$$\mu_{Z|\overline{Y}=F^*x}((\zeta_m)_{1\le m\le M})\quad\propto\quad\prod_{m=1}^M\exp(-\psi_m(\eta_m)),\tag{25}$$

where $(\eta_m)_{1\le m\le M}=F^*x$. We conclude that (23) assumes the following form.

Problem 23 Let $x=(\xi_k)_{1\le k\le K}$ denote a generic element in $\mathcal{H}$, and set

$$f_1\colon\mathcal{H}\to\mathbb{R}\colon x\mapsto\sum_{k=1}^K\phi_k(\xi_k)\tag{26}$$

and

$$f_2=\Psi\circ F^*+\iota_C\circ F^*,\tag{27}$$

where

$$\Psi\colon\mathcal{G}\to\,]-\infty,+\infty]\colon(\eta_m)_{1\le m\le M}\mapsto\sum_{m=1}^M\psi_m(\eta_m).\tag{28}$$

The objective is to minimize f_1+f_2 over $\mathcal{H}$.

Let us observe that, in view of Assumption 22iv), f_1 is a finite function in $\Gamma_0(\mathcal{H})$, whereas it follows from Assumptions 22ii) and 22v), and the surjectivity of F^* that $f_2\in\Gamma_0(\mathcal{H})$. Moreover, it follows from Proposition 14ii) that Condition (6) holds. Altogether, we obtain the following fact.

Proposition 24 *Problem 23 is a special case of Problem 4.*

C. Relevant proximity operators

In order to solve Problem 23 via Algorithm (19) and Corollary 21, it will be necessary to determine the proximity operators of the functions f_1 and f_2 in (26) and (27). As we shall see shortly, these proximity operators can often be determined in a closed form since, via Lemma 10, they can be written in terms of the proximity operators of the functions $(\phi_k)_{1\le k\le K}$ and $(\psi_m)_{1\le m\le M}$.

Let us start our discussion with the proximity operator of f_1. The choice of relevant potential functions $(\phi_k)_{1\le k\le K}$ for modeling frame coefficients in Assumption 22iv) is discussed in [13]. In particular, in order to promote a sparse representation, it has been proposed to employ proximal thresholder corresponding to non-differentiable potential functions of the form

$$\phi_k=\varrho_k+\sigma_{\Omega_k},\tag{29}$$

where $\Omega_k\subset\mathbb{R}$ is a closed bounded interval and ϱ_k satisfies $0=\varrho_k(0)\le\varrho_k\in\Gamma_0(\mathbb{R})$ and is differentiable at 0 [24]. For such potentials, the proximity operator of f_1 is obtained via the following proposition, where the soft thresholder relative to a nonempty closed interval $\Omega\subset\mathbb{R}$ is defined as

$$\operatorname{soft}_\Omega\colon\mathbb{R}\to\mathbb{R}\colon\xi\mapsto\begin{cases}\xi-\underline{\omega}, & \text{if } \xi<\underline{\omega};\\ 0, & \text{if } \xi\in\Omega;\\ \xi-\overline{\omega}, & \text{if } \xi>\overline{\omega},\end{cases}\tag{30}$$

with $\underline{\omega}=\inf\Omega$ and $\overline{\omega}=\sup\Omega$.

Proposition 25 [24] *Let f_1 be as in* (26) *and let* $\gamma\in\,]0,+\infty[$. *Then, for every* $x=(\xi_k)_{1\le k\le K}\in\mathcal{H}$,

$$\operatorname{prox}_{\gamma f_1}x=(\pi_k)_{1\le k\le K},\tag{31}$$

where

$$\pi_k=\begin{cases}\operatorname{prox}_{\gamma\varrho_k}(\operatorname{soft}_{\gamma\Omega_k}\xi_k), & \textit{if } \Omega_k\neq\varnothing;\\ \operatorname{prox}_{\gamma\varrho_k}\xi_k, & \textit{otherwise.}\end{cases}\tag{32}$$

An example of a commonly used function ϱ_k is the potential function of a generalized Gaussian distribution (see [3] for more details about this statistical model).

Example 26 [13, Example 4.4] Let $\omega_k\in\,]0,+\infty[$, let $p_k\in\,]1,+\infty[$, and let $\varrho_k=\omega_k|\cdot|^{p_k}$. Then, for every $\xi\in\mathbb{R}$, $\operatorname{prox}_{\varrho_k}\xi=\pi_k$ where π_k is the unique solution to

$$\xi=p_k\omega_k|\pi_k|^{p_k-1}\operatorname{sgn}(\pi_k)+\pi_k.$$

In particular, π_k is given by

$$\begin{cases} \xi + \dfrac{4\omega_k}{3\cdot 2^{1/3}}\Big(|\chi_k-\xi|^{1/3}-|\chi_k+\xi|^{1/3}\Big), \\ \quad \text{where } \chi_k=\sqrt{|\xi|^2+256\omega_k^3/729}, & \text{if } p_k=\dfrac{4}{3}; \\ \xi+\dfrac{9\omega_k^2\operatorname{sgn}(\xi)}{8}\left(1-\sqrt{1+\dfrac{16|\xi|}{9\omega_k^2}}\right), & \text{if } p_k=\dfrac{3}{2}; \\ \xi/(1+2\omega_k), & \text{if } p_k=2; \\ \operatorname{sgn}(\xi)\dfrac{\sqrt{1+12\omega_k|\xi|}-1}{6\omega_k}, & \text{if } p_k=3; \\ \left|\dfrac{\chi_k+\xi}{8\omega_k}\right|^{1/3}-\left|\dfrac{\chi_k-\xi}{8\omega_k}\right|^{1/3} \\ \quad \text{where } \chi_k=\sqrt{|\xi|^2+1/(27\omega_k)}, & \text{if } p_k=4. \end{cases}$$

Let us now turn our attention to the function f_2 (recall that ν is supplied by (22)).

Proposition 27 *Suppose that $C = C_1 \times \cdots \times C_M$, for some family $(C_m)_{1\le m\le M}$ of nonempty closed intervals in $\mathbb{R}$. Let $\gamma \in\,]0,+\infty[$, let $x \in \mathcal{H}$, let $F^*x = (\eta_m)_{1\le m\le M}$, and set $h = (P_{C_m}(\operatorname{prox}_{\nu\gamma\psi_m}\eta_m) - \eta_m)_{1\le m\le M}$. Then*

$$\operatorname{prox}_{\gamma f_2} x = x + \nu^{-1}(\langle h \mid e_k\rangle)_{1\le k\le K}. \tag{33}$$

Proof: See Appendix V. ■

Example 28 For a Laplacian additive noise, we have, for every $m \in \{1,\ldots,M\}$,

$$(\forall \eta\in\mathbb{R}) \quad \psi_m(\eta) = \omega_m|\eta-\zeta_m|, \tag{34}$$

where $\omega_m \in [0,+\infty[$. It follows from [13, Example 4.2] that

$$\operatorname{prox}_{\nu\gamma\psi_m}\eta_m = \zeta_m + \operatorname{soft}_{[-\nu\gamma\omega_m,\nu\gamma\omega_m]}(\eta_m-\zeta_m). \tag{35}$$

Example 29 Speckle noise is typically associated with the model

$$(\forall m\in\{1,\ldots,M\}) \quad Z_m = \overline{\Upsilon}_m(1+U_m), \tag{36}$$

where $\overline{\Upsilon}_m$ and U_m are independent random variables taking their values in $[0,+\infty[$ and $[-1,+\infty[$, respectively. Suppose that U_m has a uniform density over $[-\omega,\omega]$ with $\omega\in\,]0,1[$. Then $\psi_m = \iota_{[\zeta_m/(1+\omega),\zeta_m/(1-\omega)]}$. Consequently, we deduce that $\operatorname{prox}_{\nu\gamma\psi_m} = P_{[\zeta_m/(1+\omega),\zeta_m/(1-\omega)]}$ and that

$$P_{C_m}\circ\operatorname{prox}_{\nu\gamma\psi_m} = P_{[\zeta_m/(1+\omega),\zeta_m/(1-\omega)]\cap C_m}. \tag{37}$$

Example 30 For Poisson data, we have $z = (\zeta_m)_{1\le m\le M} \in \mathbb{N}^M$ and the components $(\overline{\eta}_m)_{1\le m\le M}$ of $\overline{y}$ must lie in $[0,+\infty[$ (when $\overline{\eta}_m = 0$, the Poisson distribution degenerates and $\zeta_m = 0$). For every $m\in\{1,\ldots,M\}$ and $\eta\in\,]0,+\infty[$, the Poisson probabilities are given by

$$\mu_{Z_m|\overline{\Upsilon}_m=\eta}(\zeta_m) = \frac{|\alpha\eta|^{\zeta_m}}{\zeta_m!}\exp(-\alpha\eta), \tag{38}$$

where $\alpha\in\,]0,+\infty[$ is a scaling parameter. Therefore, for each m, there are two alternatives. If $\zeta_m > 0$, then

$$(\forall\eta\in\mathbb{R}) \quad \psi_m(\eta) = \begin{cases} -\zeta_m\ln(\eta)+\alpha\eta, & \text{if } \eta>0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{39}$$

On the other hand, if $\zeta_m = 0$, then

$$(\forall\eta\in\mathbb{R}) \quad \psi_m(\eta) = \begin{cases} \alpha\eta, & \text{if } \eta\ge 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{40}$$

Equations (39) and (40) correspond to the potential functions of a gamma distribution and of an exponential distribution, respectively. We deduce from [13, Examples 4.8 and 4.9] that

$$\operatorname{prox}_{\nu\gamma\psi_m}\eta_m = \frac{\eta_m-\nu\gamma\alpha+\sqrt{|\eta_m-\nu\gamma\alpha|^2+4\nu\gamma\zeta_m}}{2}. \tag{41}$$

VI. NUMERICAL EXAMPLES

We consider the problem of denoising images of size $N\times N$, where $N = 512$. In our experiments, the distribution of the noise is assumed to be known.

Using the notation of Section V, the dimension of the signal space $\mathcal{G}$ is $M = N^2$. The size of the frame is set to $K = 4M$ and $(e_k)_{1\le k\le K}$ is chosen to be a two-dimensional wavelet frame. More precisely, we use the concatenation of four shifted separable dyadic orthonormal wavelet decompositions [41] carried out over 4 resolution levels. The shift parameters are $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$ and symlet filters of length 8 are used [26]. Denoising is performed by solving Problem 23, where f_1 is the nondifferentiable function given by (26), (29), and Example 26, with $p_k\in\{4/3,3/2,2\}$ and $\Omega_k = [-\tau_k,\tau_k]$ with $\tau_k\in\,]0,+\infty[$. For each subband, an adapted value of (τ_k,ω_k,p_k) is selected via a maximum likelihood approach. On the other hand, the function f_2 is given by (27) where the constraint set is $C = [0,255]^M$ and where the functions $(\psi_m)_{1\le m\le M}$ defining Ψ in (28) are chosen according to the noise distribution. In the two scenarios investigated below, we have $\inf\Psi(\mathcal{H}) > -\infty$. Hence, as a result of the coercivity of the functions $(\phi_k)_{1\le k\le K}$ and [13, Proposition 3.3(iii)(c)], f_1+f_2 is coercive and Problem 23 admits at least one solution. The denoised images are obtained via Algorithm (19). The convergence of the sequence $(x_{n+\frac{1}{2}})_{n\in\mathbb{N}}$ to a solution to Problem 23 is guaranteed by Corollary 21 and Proposition 24.

As a measure of discrepancy between two images y_1 and y_2, we subsequently employ the relative error

$$\Delta(y_1,y_2) = 20\log_{10}(\|y_2\|/\|y_1-y_2\|). \tag{42}$$

Let us point out that, for frame-based denoising problems, the Bayesian formulation adopted in [13] requires stronger assumptions than those made in this paper, while allowing for the use of arbitrary frame representations. In particular, the noise distributions considered in the following examples result in a nondifferentiable function f_2 and they can therefore not be dealt with via the method developed in [13], which is based on Theorem 3.

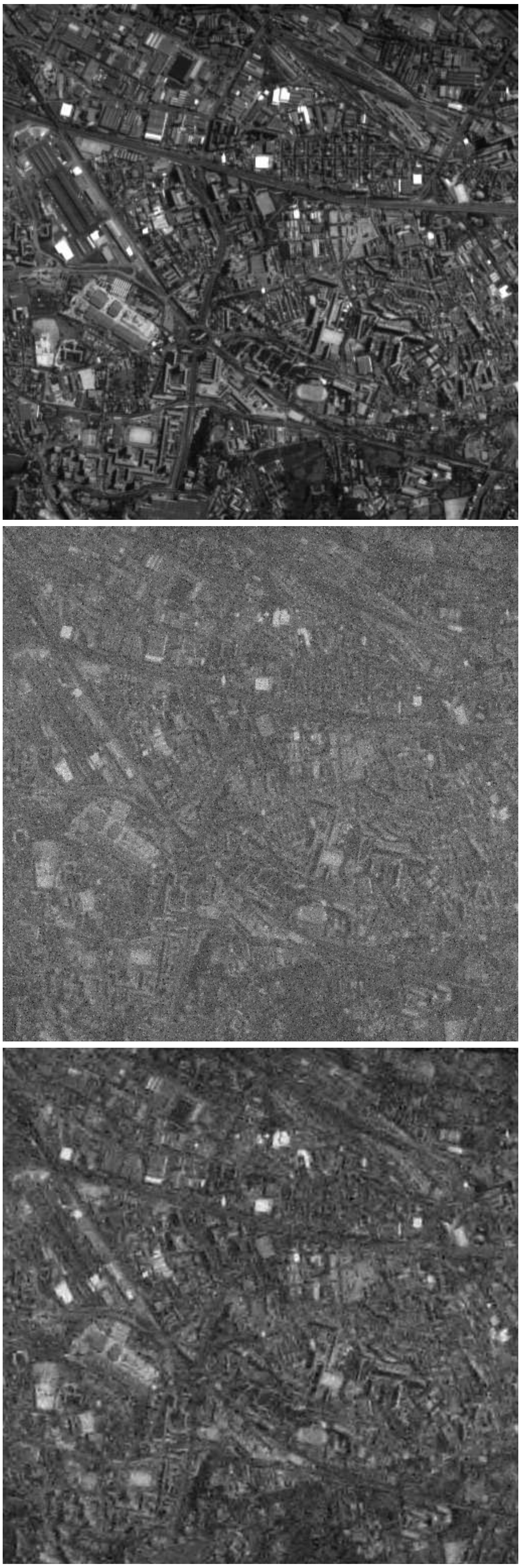


Fig. 1. Example 1. Top: original image; center: noisy image (for improved readability the Laplacian noise realizations have been limited to the interval $[-200, 200]$, which captures 99.9% of the samples); bottom: denoised image.

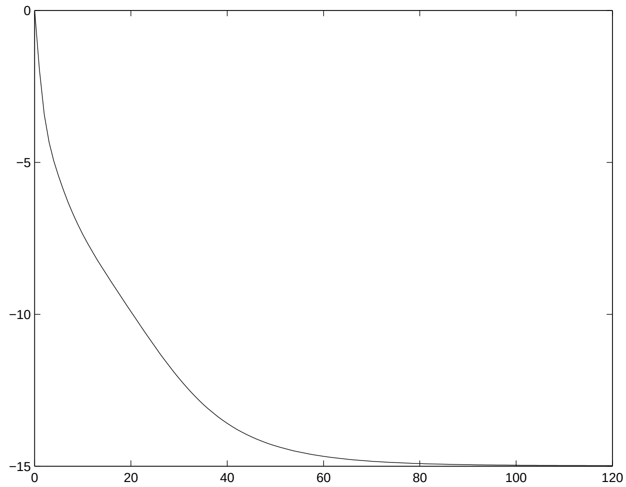


Fig. 2. Example 1. $(f_1(x_{n+1/2}) + f_2(x_{n+1/2}))/(f_1(x_{1/2}) + f_2(x_{1/2}))$ in dB versus the iteration index n with $\gamma = 50$ and $\lambda_n \equiv 1$.

A. Example 1: Laplacian noise

The original image $\overline{y}$ shown in Fig. 1 (top) has been corrupted by addition of i.i.d. zero-mean Laplacian noise. The degraded image z can be seen in Fig. 1 (center). The image-to-noise ratio is $\Delta(z, \overline{y}) = 5.95$ dB. The functions $(\psi_m)_{1\leq m\leq M}$ are as in Example 28. The denoised image $\tilde{y}$ is shown in Fig. 1 (bottom). The relative mean square error with respect to the original image is $\Delta(\tilde{y}, \overline{y}) = 14.56$ dB. The values taken by the objective function along the iterations are plotted in Fig. 2, showing the good asymptotic behavior of Algorithm (19).

B. Example 2: Poisson noise

The original image $\overline{y}$ is displayed in Fig. 3 (top). It is used to generate the Poisson field z shown in Fig. 3 (center) via the probabilistic model (38), where $\alpha = 0.10$. The relative error with respect to the scaled original image is $\Delta(z, \alpha\overline{y}) = 11.66$ dB. The functions $(\psi_m)_{1\leq m\leq M}$ are as in Example 30. The restored image $\tilde{y}$ shown in Fig. 3 (bottom) yields a relative error of $\Delta(\tilde{y}, \overline{y}) = 21.02$ dB. The convergence pattern of the algorithm is illustrated in Fig. 4. Note that, in contrast with the previous example, the objective function is now negative valued.

VII. Closing remarks

We have proposed an algorithm for solving signal recovery problems that can be modeled via the minimization of the sum of two lower semicontinuous convex functions in a Hilbert space. In this variational framework, neither of the functions need be differentiable or finite and, thereby, a broad class of signal recovery problems is captured. On an algorithmic level, the proposed method (19) is derived from the Douglas-Rachford splitting algorithm for finding zeros of monotone operators. Its main advantage is to employ the functions separately through individual proximity operators. This makes the implementation much easier than in alternative nonsmooth optimization methods such as (7), which are tailored for generic minimization problems and in which the sum of the

Fig. 3. Example 2. Top: original image; center: noisy image; bottom: denoised image.

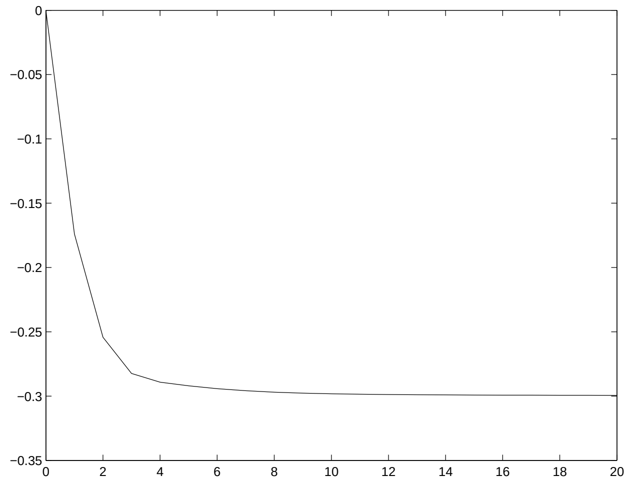


Fig. 4. Example 2. $(f_1(x_{1/2}) + f_2(x_{1/2}))/(f_1(x_{n+1/2}) + f_2(x_{n+1/2}))$ in dB versus the iteration index n with $\gamma = 140$ and $\lambda_n \equiv 1$.

two functions has to be dealt with directly. We have shown (Theorem 20) that the proposed Douglas-Rachford algorithm does produce a solution to the variational problem regardless of the starting point.

Each iteration of algorithm (19) requires the computation of the proximity operators of both functions separately. Although many functions encountered in signal recovery lead to closed-form proximity operators [13], [24], [25], the implementation of such operators may be costly in general as they amount to solving (strongly convex) minimization subproblems. It would therefore be worthwhile to attempt to replace these operators with approximations that would be less demanding numerically, while preserving the convergence properties of the algorithm. Another limitation of our algorithm is that it is inherently structured for the sum of two functions. Variants that could handle efficiently the sum of more than two nondifferentiable functions should be explored for greater flexibility.

APPENDIX I
PROOF OF PROPOSITION 11

Set $g = f \circ L$. It follows from the linearity of L and the convexity of f that g is convex. In addition, the continuity of L and the lower semicontinuity of f imply that g is lower semicontinuous. Finally, since (15) implies that $0 \in \operatorname{dom} f - \operatorname{ran} L$, we have $\operatorname{dom} f \cap \operatorname{ran} L \neq \varnothing$ and, in turn, $\operatorname{dom} g \neq \varnothing$. Therefore $g \in \Gamma_0(\mathcal{H})$.

To prove the second assertion, fix $x \in \mathcal{H}$ and set $p = \operatorname{prox}_g x$. It follows from (15) and [57, Theorem 2.8.3] that $\partial g = L^* \circ \partial f \circ L$. Hence, Lemma 7 yields

$$\begin{aligned} p = \operatorname{prox}_g x \quad &\Leftrightarrow \quad x - p \in \partial g(p) = L^*(\partial f(Lp)) && (43)\\ &\Rightarrow \quad L(x-p) \in L(L^*(\partial f(Lp))) \\ &\Leftrightarrow \quad Lx - Lp \in \nu \partial f(Lp) = \partial(\nu f)(Lp) \\ &\Leftrightarrow \quad Lp = \operatorname{prox}_{\nu f}(Lx). && (44) \end{aligned}$$

Now set $V = \ker L$. It follows from [28, Chapter 8] that

$V^\perp = \operatorname{ran} L^*$ and that

$$\begin{cases} P_V = \operatorname{Id} - (L^* \circ (L \circ L^*)^{-1} \circ L) = \operatorname{Id} - \nu^{-1} L^* \circ L \\ P_{V^\perp} = \operatorname{Id} - P_V = \nu^{-1} L^* \circ L. \end{cases} \tag{45}$$

Hence, since (43) implies that $p - x \in \operatorname{ran} L^*$, we obtain

$$P_V p = P_V x + P_V(p - x) = P_V x = (\operatorname{Id} - \nu^{-1} L^* \circ L)x. \tag{46}$$

On the other hand, we derive from (45) and (44) that

$$P_{V^\perp} p = \nu^{-1} L^*(Lp) = \nu^{-1}(L^* \circ \operatorname{prox}_{\nu f} \circ L)x. \tag{47}$$

Altogether,

$$p = P_V p + P_{V^\perp} p = x + \nu^{-1}(L^* \circ (\operatorname{prox}_{\nu f} - \operatorname{Id}) \circ L)x, \tag{48}$$

and we obtain (16). ∎

Appendix II
Proof of Proposition 12

Fix $x \in \mathcal{H}$, and set $g \colon y \mapsto f(y) + \|x - y\|^2/2$ and $p = \operatorname{prox}_f x$.

i): By definition, p minimizes g. Therefore, if p belongs to C, it also minimizes $\iota_C + g$, whence $p = \operatorname{prox}_{\iota_C + f} x$.

ii): Since C is a closed interval in $\mathbb{R}$, we must show that

$$\operatorname{prox}_{f + \iota_C} x = \begin{cases} \inf C, & \text{if } p < \inf C; \\ p, & \text{if } p \in C; \\ \sup C, & \text{if } p > \sup C. \end{cases} \tag{49}$$

If $p \in C$, the identity follows from i). Now suppose that $p < \inf C$. Since g is strictly convex and admits p as its unique minimizer, it increases strictly over the interval $[p, +\infty[\cap \operatorname{dom} f$. Since $C \cap \operatorname{dom} g \neq \varnothing$, $\inf C \in C \cap \operatorname{dom} g$ and $\inf C$ therefore minimizes $\iota_C + g$, which shows that $\operatorname{prox}_{\iota_C + f} x = \inf C$. The case when $p > \sup C$ is treated analogously. ∎

Appendix III
Proof of Proposition 14

i): The assumption implies that $0 \in \operatorname{int}(\operatorname{dom} f_1 - \operatorname{dom} f_2)$. Therefore $\operatorname{cone}(\operatorname{dom} f_1 - \operatorname{dom} f_2) = \mathcal{H}$ and hence (6) holds.

ii): If f_1 is finite, then $\operatorname{dom} f_1 = \mathcal{H}$ and we obtain a special case of i).

iii): The assumption implies that $0 \in (\operatorname{rint} \operatorname{dom} f_1 - \operatorname{rint} \operatorname{dom} f_2)$. However, since f_1 and f_2 are convex functions, the sets $\operatorname{dom} f_1$ and $\operatorname{dom} f_2$ are convex and we derive from [42, Corollary 6.6.2] that $\operatorname{rint} \operatorname{dom} f_1 - \operatorname{rint} \operatorname{dom} f_2 = \operatorname{rint}(\operatorname{dom} f_1 - \operatorname{dom} f_2)$. Hence, $0 \in \operatorname{rint}(\operatorname{dom} f_1 - \operatorname{dom} f_2)$ and, in turn, $\operatorname{cone}(\operatorname{dom} f_1 - \operatorname{dom} f_2)$ is the span of $\operatorname{dom} f_1 - \operatorname{dom} f_2$, which is closed. ∎

Appendix IV
Proof of Proposition 18

i): It results from the assumptions on f_1 and f_2 that $f_1 + f_2$ lies in $\Gamma_0(\mathcal{H})$ and that it is coercive. Hence, the claim follows from [57, Theorem 2.5.1(ii)].

ii): See [57, Proposition 2.5.6].

iii): It follows from (6) and [57, Theorem 2.8.3] that $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$. Hence, we derive from Lemma 6 that a) $\Leftrightarrow$ $x \in \operatorname{zer}(\partial(f_1 + f_2)) = \operatorname{zer}(\partial f_1 + \partial f_2) \Leftrightarrow$ b). Finally, the equivalence b) $\Leftrightarrow$ c) follows from Lemma 5, Lemma 8, and [21, Lemma 2.6]. ∎

Appendix V
Proof of Proposition 27

We first deduce from (28) and Assumption 22v) that $\operatorname{dom}(\iota_C + \gamma\Psi) = \operatorname{dom} \iota_C \cap \operatorname{dom} \Psi = C \cap (\operatorname{dom} \psi_1 \times \cdots \times \operatorname{dom} \psi_M) \neq \varnothing$. Hence, since F^* is surjective, $\operatorname{ran} F^* = \mathcal{G}$ and therefore

$$\operatorname{cone}(\operatorname{dom}(\gamma\Psi + \iota_C) - \operatorname{ran} F^*) = \mathcal{G}. \tag{50}$$

In turn, it follows from (20), (22), (27), and Proposition 11 with $L = F^*$ and $f = \iota_C + \gamma\Psi$ that, for every $x \in \mathcal{H}$,

$$\begin{aligned} \operatorname{prox}_{\gamma f_2} x &= x + \nu^{-1} F(\operatorname{prox}_{\iota_C + \nu\gamma\Psi}(F^* x) - F^* x) \\ &= x + \nu^{-1}(\langle p - y \mid e_k \rangle)_{1 \le k \le K}, \end{aligned} \tag{51}$$

where $y = F^* x$ and $p = \operatorname{prox}_{\iota_C + \nu\gamma\Psi} y$. We also note that

$$\iota_C(y) + \nu\gamma\Psi(y) = \sum_{m=1}^{M} \iota_{C_m}(\eta_m) + \nu\gamma\psi_m(\eta_m). \tag{52}$$

On the other hand, it follows from Proposition 12ii) and Assumption 22 that, for every $m \in \{1, \dots, M\}$,

$$\operatorname{prox}_{\iota_{C_m} + \nu\gamma\psi_m} = P_{C_m} \circ \operatorname{prox}_{\nu\gamma\psi_m}. \tag{53}$$

Setting $\mathbb{K} = \{1, \dots, M\}$ and letting $(o_m)_{1 \le m \le M}$ be the canonical basis of $\mathcal{G}$ in Lemma 10, we therefore obtain

$$p = (P_{C_m}(\operatorname{prox}_{\nu\gamma\psi_m} \eta_m))_{1 \le m \le M}. \tag{54}$$

Combining this identity with (51) yields (33). ∎

References

[1] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Processing,* vol. 14, pp. 1860–1875, 2005.

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

[3] A. Antoniadis, D. Leporini and J.-C. Pesquet, "Wavelet thresholding for some classes of non-Gaussian noise," *Statist. Neerlandica,* vol. 56, pp. 434–453, 2002.

[4] J.-P. Aubin, *Optima and Equilibria – An Introduction to Nonlinear Analysis,* 2nd ed. New York: Springer-Verlag, 1998.

[5] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis.* Boston, MA: Birkhäuser, 1990.

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

[7] H. H. Bauschke, P. L. Combettes, and S. G. Kruk, "Extrapolation algorithm for affine-convex feasibility problems," *Numer. Algorithms,* vol. 41, pp. 239–274, 2006.

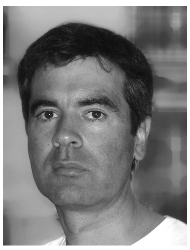
[8] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Finding best approximation pairs relative to two closed convex sets in Hilbert spaces," *J. Approx. Theory,* vol. 127, pp. 178–192, 2004.

[9] P. Besbeas, I. De Feis, and T. Sapatinas, "A comparative simulation study of wavelet shrinkage estimators for Poisson counts," *Int. Statist. Rev.,* vol. 72, pp. 209–237, 2004.

[10] J. M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: A GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Trans. Image Processing,* vol. 15, pp. 937–951, 2006.

[11] E. J. Candès and D. L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *Ann. Statist.,* vol. 30, pp. 784–842, 2002.

[12] Y. Censor, T. Bortfeld, B. Martin, and A. Tromov, "A unified approach for inversion problems in intensity-modulated radiation therapy," *Phys. Med. Biol.,* vol. 51, pp. 2353–2365, 2006.
[13] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems,* vol. 23, pp. 1495–1518, 2007.
[14] C. Chaux, L. Duval, and J.-C. Pesquet, "Image analysis using a dual-tree M-band wavelet transform," *IEEE Trans. Image Processing,* vol. 15, pp. 2397–2412, 2006.
[15] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE,* vol. 81, pp. 182–208, 1993.
[16] P. L. Combettes, "Inconsistent signal feasibility problems: Least-squares solutions in a product space," *IEEE Trans. Signal Processing,* vol. 42, pp. 2955–2966, 1994.
[17] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics* (P. Hawkes, Ed.), vol. 95, pp. 155–270. New York: Academic Press, 1996.
[18] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Trans. Image Processing,* vol. 6, pp. 493–506, 1997.
[19] P. L. Combettes, "Hilbertian convex feasibility problem: Convergence of projection methods," *Appl. Math. Optim.,* vol. 35, pp. 311–330, 1997.
[20] P. L. Combettes, "Convexité et signal," Proc. Congrès de Mathématiques Appliquées et Industrielles SMAI'01, pp. 6–16. Pompadour, France, May 28–June 1, 2001.
[21] P. L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization,* vol. 53, pp. 475–504, 2004.
[22] P. L. Combettes and J. Luo, "An adaptive level set method for nondifferentiable constrained image recovery," *IEEE Trans. Image Processing,* vol. 11, pp. 1295–1304, 2002.
[23] P. L. Combettes and J.-C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Trans. Image Processing,* vol. 13, pp. 1213–1222, 2004.
[24] P. L. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM J. Optim.,* to appear.
[25] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.,* vol. 4, pp. 1168–1200, 2005.
[26] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[27] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.
[28] F. Deutsch, *Best Approximation in Inner Product Spaces*. New York: Springer-Verlag, 2001.
[29] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Processing,* vol. 14, pp. 2091–2106, 2005.
[30] J. Douglas and H. H. Rachford, "On the numerical solution of heat conduction problems in two or three space variables," *Trans. Amer. Math. Soc.*, vol. 82, pp. 421–439, 1956.
[31] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programming,* vol. 55, pp. 293–318, 1992.
[32] B. Eicke, "Iteration methods for convexly constrained ill-posed problems in Hilbert space," *Numer. Funct. Anal. Optim.,* vol. 13, pp. 413–429, 1992.
[33] M. Elad, "Why simple shrinkage is still relevant for redundant representations?," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5559–5569, 2006.
[34] X. Fan, H. Huang, D. Liang, and C. Qi, "A hybrid parallel projection approach to object-based image restoration," *Pattern Recog. Lett.,* vol. 27, pp. 1045-1053, 2006.
[35] F. J. González-Castaño, U. M. García-Palomares, J. L. Alba-Castro, and J. M. Pousada-Carballo, "Fast image recovery using dynamic load balancing in parallel architectures, by means of incomplete projections," *IEEE Trans. Image Processing,* vol. 10, pp. 493–499, 2001.
[36] D. Han and D. R. Larson, *Frames, Bases, and Group Representations,* Mem. Amer. Math. Soc., vol. 147. Providence, RI: 2000.
[37] G. T. Herman, *Image Reconstruction from Projections, the Fundamentals of Computerized Tomography.* New York: Academic Press, 1980.
[38] P.-L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.
[39] J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Sér. A Math.,* vol. 255, pp. 2897–2899, 1962.
[40] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France,* vol. 93, pp. 273–299, 1965.
[41] J.-C. Pesquet, H. Krim, and H. Carfantan, "Time-invariant orthonormal wavelet representations," *IEEE Trans. Signal Processing*, vol. 44, pp. 1964–1970, 1996.
[42] R. T. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton University Press, 1970.
[43] A. A. Samsonov, E. G. Kholmovski, D. L. Parker, and C. R. Johnson, "POCSENSE: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging," *Magnet. Reson. Med.,* vol. 52, pp. 1397–1406, 2004.
[44] S. Sardy, A. Antoniadis, and P. Tseng, "Automatic smoothing with wavelets for a wide class of distributions," *J. Comput. Graph. Statist.*, vol. 13, pp. 399–421, 2004.
[45] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions.* New York: Springer-Verlag, 1985.
[46] H. Stark (Editor), *Image Recovery: Theory and Application.* San Diego, CA: Academic Press, 1987.
[47] H. Stark and Y. Yang, *Vector Space Projections : A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics.* New York: Wiley, 1998.
[48] A. M. Thompson and J. Kay, "On some Bayesian choices of regularization parameter in image restoration," *Inverse Problems,* vol. 9, pp. 749–761, 1993.
[49] R. Tolimieri and M. An, *Time-Frequency Representations.* Boston, MA: Birkhäuser, 1998.
[50] H. J. Trussell, "*A priori* knowledge in algebraic reconstruction methods," in *Advances in Computer Vision and Image Processing* (T. S. Huang, Ed.), vol. 1, pp. 265–316. Greenwich, CT: JAI Press, 1984.
[51] R. S. Varga, *Matrix Iterative Analysis*, 2nd edition. New York: Springer-Verlag, 2000.
[52] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," *IEEE Trans. Signal Processing,* vol. 50, pp. 1091–1101, 2002.
[53] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Trans. Circuits Syst.,* vol. 25, pp. 694–702, 1978.
[54] D. C. Youla and V. Velasco, "Extensions of a result on the synthesis of signals in the presence of inconsistent constraints," *IEEE Trans. Circuits Syst.,* vol. 33, pp. 465–468, 1986.
[55] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 – theory," *IEEE Trans. Medical Imaging,* vol. 1, pp. 81–94, 1982.
[56] M. Yukawa and I. Yamada, "Pairwise optimal weight realization – Acceleration technique for set-theoretic adaptive parallel subgradient projection algorithm," *IEEE Trans. Signal Processing,* vol. 54, pp. 4557–4571, 2006.
[57] C. Zălinescu, *Convex Analysis in General Vector Spaces.* River Edge, NJ: World Scientific, 2002.
[58] E. Zeidler, *Nonlinear Functional Analysis and Its Applications,* vols. I-V. New York: Springer-Verlag, 1985–1990.



Patrick Louis Combettes (S'84-M'90-SM'96-F'06) is a Professor with the Laboratoire Jacques-Louis Lions, Faculté de Mathématiques, Université Pierre et Marie Curie – Paris 6, Paris, France.
Dr. Combettes is recipient of the IEEE Signal Processing Society Paper Award.

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

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