

On a Decomposition Formula for the Proximal Operator of the Sum of Two Convex Functions

Samir Adly

*Institut de Recherche XLIM, Université de Limoges,
123 Avenue A. Thomas, 87060 Limoges, France
samir.adly@unilim.fr*

Loïc Bourdin

*Institut de Recherche XLIM, Université de Limoges,
123 Avenue A. Thomas, 87060 Limoges, France
loic.bourdin@unilim.fr*

Fabien Caubet

*Institut de Mathématiques de Toulouse, Université de Toulouse,
118 Route de Narbonne 31062 Toulouse
fabien.caubet@math.univ-toulouse.fr*

Received: February 21, 2018

Revised manuscript received: June 7, 2018

Accepted: June 11, 2018

The main result of the present theoretical paper is an original decomposition formula for the proximal operator of the sum of two proper, lower semicontinuous and convex functions f and g . For this purpose, we introduce a new operator, called f -proximal operator of g and denoted by prox_g^f , that generalizes the classical notion. Then we prove the decomposition formula $\text{prox}_{f+g} = \text{prox}_f \circ \text{prox}_g^f$. After collecting several properties and characterizations of prox_g^f , we prove that it coincides with the fixed points of a generalized version of the classical Douglas-Rachford operator. This relationship is used for the construction of a weakly convergent algorithm that computes numerically this new operator prox_g^f , and thus, from the decomposition formula, allows to compute numerically prox_{f+g} . It turns out that this algorithm was already considered and implemented in previous works, showing that prox_g^f is already present (in a hidden form) and useful for numerical purposes in the existing literature. However, to the best of our knowledge, it has never been explicitly expressed in a closed formula and neither been deeply studied from a theoretical point of view. The present paper contributes to fill this gap in the literature. Finally we give an illustration of the usefulness of the decomposition formula in the context of sensitivity analysis of linear variational inequalities of second kind in a Hilbert space.

Keywords: Convex analysis, proximal operator, Douglas-Rachford operator, Forward-Backward operator.

2010 Mathematics Subject Classification: 46N10, 47N10, 49J40, 49Q12.

1. Introduction, notations and basics

1.1. Introduction

The *proximal operator* (also known as *proximity operator*) of a proper, lower semicontinuous, convex and extended-real-valued function was first introduced

by Moreau in 1962 in [11, 12] and can be viewed as an extension of the projection operator on a nonempty closed and convex subset of a Hilbert space. This wonderful tool plays an important role, from both theoretical and numerical points of view, in applied mathematics and engineering sciences. This paper fits within the wide theoretical literature dealing with the proximal operator. For the rest of this introduction, we use standard notations of convex analysis. For the reader who is not acquainted with convex analysis, we refer to Section 1.2 for notations and basics.

Motivations from a sensitivity analysis. The present work was initially motivated by the sensitivity analysis, with respect to a nonnegative parameter $t \geq 0$, of a parameterized linear variational inequality of second kind in a Hilbert space H , with a corresponding function $h \in \Gamma_0(H)$, where $\Gamma_0(H)$ is the set of proper, lower semicontinuous and convex functions from H into $\mathbb{R} \cup \{+\infty\}$. More precisely, for all $t \geq 0$, we consider the problem of finding $u(t) \in H$ such that

$$\langle u(t), z - u(t) \rangle + h(z) - h(u(t)) \geq \langle r(t), z - u(t) \rangle, \quad (1)$$

for all $z \in H$, where $r: \mathbb{R}^+ \rightarrow H$ is assumed to be given and smooth enough. In that framework, the solution $u(t) \in H$ (which depends on the parameter t) can be expressed in terms of the proximal operator of h denoted by prox_h . Precisely it holds that $u(t) = \text{prox}_h(r(t))$ for all $t \geq 0$. As a consequence, the differentiability of $u(\cdot)$ at $t = 0$ is strongly related to the regularity of prox_h . If h is a smooth function, one can easily compute (from the classical inverse mapping theorem for instance) the differential of prox_h , and then the sensitivity analysis can be achieved. In that smooth case, note that the variational inequality (1) can actually be reduced to an equality. On the other hand, if $h = \iota_K$ is the indicator function of a nonempty closed and convex subset $K \subset H$, then $\text{prox}_h = \text{proj}_K$ is the classical projection operator on K . In that case, the work of Mignot in [10, Theorem 2.1 p.145] (see also the work of Haraux in [8, Theorem 2 p.620]) provides an asymptotic expansion of $\text{prox}_h = \text{proj}_K$ and permits to obtain a differentiability result on $u(\cdot)$ at $t = 0$.

In a parallel work (in progress) of the authors on some shape optimization problems with unilateral contact and friction, the considered variational inequality (1) involves the sum of two functions. Precisely, $h = f + g$ where $f = \iota_K$ (K being a nonempty closed and convex set of constraints), and where $g \in \Gamma_0(H)$ is a smooth function (derived from the regularization of the friction functional in view of a numerical treatment). Despite the regularity of g , note that the variational inequality (1) cannot be reduced to an equality due to the presence of the constraint set K . In that framework, in order to get an asymptotic expansion of $\text{prox}_h = \text{prox}_{f+g}$, a first and natural strategy would be to look for a convenient explicit expression of prox_{f+g} in terms of prox_f and prox_g . Unfortunately, this theoretical question still remains an open challenge in the literature. Let us mention that Yu provides in [18] some necessary and/or sufficient conditions on general functions $f, g \in \Gamma_0(H)$ under which $\text{prox}_{f+g} = \text{prox}_f \circ \text{prox}_g$. Unfortunately, as underlined by the author himself, these conditions are very restrictive and are not satisfied in most of cases (see, e.g., [18, Example 2] for a counterexample).

Before coming to the main topic of this paper, we recall that a wide literature is already concerned with the sensitivity analysis of parameterized (linear and nonlinear) variational inequalities. We refer for instance to [3, 8, 13, 17] and references therein. The results in are considered in very general frameworks. We precise that our original objective was to look for a simple and compact formula for the derivative $u'(0)$ in the very particular case described above, that is, in the context of a linear variational inequality and with $h = f + g$ where f is an indicator function and g is a smooth function. For this purpose, we were led to consider the proximal operator of the sum of two functions in $\Gamma_0(H)$, to introduce a new operator and finally to prove the results presented in this paper.

Introduction of the f -proximal operator and main result. Let us consider general functions $f, g \in \Gamma_0(H)$. In order to avoid trivialities, we will assume in the whole paper that $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ when dealing with the sum $f + g$.

Section 2 is devoted to the introduction (see Definition 2.1) of a new operator $\text{prox}_g^f: H \rightrightarrows H$ called *f -proximal operator of g* and defined by

$$\text{prox}_g^f := (\text{I} + \partial g \circ \text{prox}_f)^{-1}. \quad (2)$$

This new operator can be seen as a generalization of prox_g in the sense that, if f is constant for instance, then $\text{prox}_g^f = \text{prox}_g$. More general sufficient (and necessary) conditions under which $\text{prox}_g^f = \text{prox}_g$ are provided in Propositions 2.13 and 2.15. We prove in Proposition 2.5 that the domain of prox_g^f satisfies $D(\text{prox}_g^f) = H$ if and only if $\partial(f + g) = \partial f + \partial g$. Note that prox_g^f is a set-valued operator *a priori*. We provide in Proposition 2.18 some sufficient conditions under which prox_g^f is single-valued. Some examples illustrate all the previous results throughout the section (see Examples 2.2, 2.3, 2.4, 2.7 and 2.17).

Finally, if the additivity condition $\partial(f + g) = \partial f + \partial g$ is satisfied, the main result of the present paper (see Theorem 2.8) is the original decomposition formula

$$\text{prox}_{f+g} = \text{prox}_f \circ \text{prox}_g^f. \quad (3)$$

It is well-known in the literature that obtaining a theoretical formula for prox_{f+g} is not an easy task in general, even if prox_f and prox_g are known. We give a more precise description of the difficulty to obtain an easy computable formula of prox_{f+g} in Appendix A, which claims that there is no closed formula, independent of f and g , allowing to write prox_{f+g} as a linear combination of compositions of linear combinations of I , prox_f , prox_g , prox_f^{-1} and prox_g^{-1} . In the decomposition formula (3), it should be noted that the difficulty of computing prox_{f+g} is only transferred to the computation of prox_g^f which is not an easier task. Note that other rewritings, which are not suitable for an easy computation of prox_{f+g} neither, can be considered such as

$$\text{prox}_{f+g} = (\text{prox}_f^{-1} + \text{prox}_g^{-1} - \text{I})^{-1} = (\text{prox}_{2f}^{-1} + \text{prox}_{2g}^{-1})^{-1} \circ 2\text{I},$$

the second equality being provided in [2, Corollary 25.35 p.458]. However we show in this paper that our decomposition formula (3) is of theoretical interest in

order to prove in a concise and elegant way almost all other new statements of this paper, and also to recover in a simple way some well-known results (see Sections 3.3 and 4.1 for instance), making it central in our work. We provide an illustration of this feature in the next paragraph about the classical Douglas-Rachford algorithm. Moreover, as explained in the last paragraph of this introduction, we also prove in this paper the usefulness of the decomposition formula (3) in the context of sensitivity analysis of the variational inequality (1) (see Section 4.2).

Relationship with the classical Douglas-Rachford operator. Recall that the proximal operator prox_{f+g} is strongly related to the minimization problem

$$\operatorname{argmin} f + g,$$

since the set of solutions is exactly the set of fixed points of prox_{f+g} denoted by $\text{Fix}(\text{prox}_{f+g})$. In the sequel, we will assume that the above minimization problem admits at least one solution. The classical *Douglas-Rachford operator*, introduced in [6] and denoted here by $\mathcal{DR}_{f,g}$ (see Section 3 for details), provides an algorithm $x_{n+1} = \mathcal{DR}_{f,g}(x_n)$ that is weakly convergent to some $x^* \in \mathbb{H}$ satisfying

$$\text{prox}_f(x^*) \in \operatorname{argmin} f + g.$$

Even if the *Douglas-Rachford algorithm* is not a *proximal point algorithm* in general, in the sense that $\mathcal{DR}_{f,g}$ is not equal to prox_φ for some $\varphi \in \Gamma_0(\mathbb{H})$ in general, it is a very powerful tool since it allows to solve the above minimization problem, requiring only the knowledge of prox_f and prox_g . We refer to [2, Section 28.3 p.517] for more details.

Section 3 deals with the relations between the Douglas-Rachford operator $\mathcal{DR}_{f,g}$ and the f -proximal operator prox_g^f introduced in this paper. Precisely, we prove in Proposition 3.2 that, for all $x \in \mathbb{H}$,

$$\text{prox}_g^f(x) = \text{Fix}(\overline{\mathcal{DR}}_{f,g}(x, \cdot)),$$

where $\overline{\mathcal{DR}}_{f,g}(x, \cdot)$ denotes an x -dependent generalization of the classical Douglas-Rachford operator $\mathcal{DR}_{f,g}$, in the sense that $\mathcal{DR}_{f,g}(y) = \overline{\mathcal{DR}}_{f,g}(\text{prox}_f(y), y)$ for all $y \in \mathbb{H}$. We refer to Section 3 for the precise definition of $\overline{\mathcal{DR}}_{f,g}(x, \cdot)$ that only depends on the knowledge of prox_f and prox_g .

Let us show that the above statements, in particular the decomposition formula (3), allow to recover in a concise way the well-known inclusion

$$\text{prox}_f(\text{Fix}(\mathcal{DR}_{f,g})) \subset \operatorname{argmin} f + g = \text{Fix}(\text{prox}_{f+g}). \quad (4)$$

Indeed, if $x^* \in \text{Fix}(\mathcal{DR}_{f,g})$, then $x^* \in \text{Fix}(\overline{\mathcal{DR}}_{f,g}(\text{prox}_f(x^*), \cdot)) = \text{prox}_g^f(\text{prox}_f(x^*))$. From the decomposition formula (3), we conclude that

$$\text{prox}_f(x^*) = \text{prox}_f \circ \text{prox}_g^f(\text{prox}_f(x^*)) = \text{prox}_{f+g}(\text{prox}_f(x^*)).$$

This proof of only few lines is an illustration of the theoretical interest of the decomposition formula (3). Note that the above inclusion (4) is, as well-known, an equality (see Section 3.3 and Proposition 3.8 for details).

The f -proximal operator prox_g^f introduced in this paper is also of interest from a numerical point of view. Indeed, if $x \in \text{D}(\text{prox}_g^f)$, we prove in Theorem 3.3 that the fixed-point algorithm $y_{k+1} = \overline{\mathcal{DR}}_{f,g}(x, y_k)$, denoted by (\mathcal{A}_1) , weakly converges to some $y^* \in \text{prox}_g^f(x)$. Moreover, if the additivity condition $\partial(f+g) = \partial f + \partial g$ is satisfied, we get from the decomposition formula (3) that $\text{prox}_f(y^*) = \text{prox}_{f+g}(x)$. In that situation, we conclude that Algorithm (\mathcal{A}_1) allows to compute numerically $\text{prox}_{f+g}(x)$ with the only knowledge of prox_f and prox_g . It turns out that Algorithm (\mathcal{A}_1) was already considered, up to some translations, and implemented in previous works (see, e.g., [4, Algorithm 3.5]), showing that the f -proximal operator prox_g^f is already present (in a hidden form) and useful for numerical purposes in the existing literature. However, to the best of our knowledge, it has never been explicitly expressed in a closed formula such as (2) and neither been deeply studied from a theoretical point of view. The present paper contributes to fill this gap in the literature.

Some other applications and forthcoming works. Section 4 can be seen as a conclusion of the paper. Its aim is to provide a glimpse of some other applications of our main result (Theorem 2.8) and to raise open questions for forthcoming works. This section is splitted into two parts.

In Section 4.1 we consider the framework where $f, g \in \Gamma_0(\mathbb{H})$ with g differentiable on \mathbb{H} . In that framework, we prove from the decomposition formula (3) that prox_{f+g} is related to the classical *Forward-Backward operator* (see [5, Section 10.3 p.191] for details) denoted by $\mathcal{FB}_{f,g}$. Precisely, we prove in Proposition 4.1 that

$$\text{prox}_{f+g}(x) = \text{Fix}(\overline{\mathcal{FB}}_{f,g}(x, \cdot)),$$

for all $x \in \mathbb{H}$, where $\overline{\mathcal{FB}}_{f,g}(x, \cdot)$ denotes a x -dependent generalization of the classical Forward-Backward operator $\mathcal{FB}_{f,g}$. We refer to Section 4.1 for the precise definition of $\overline{\mathcal{FB}}_{f,g}(x, \cdot)$ that only depends on the knowledge of prox_f and ∇g . From this point, one can develop a similar strategy as in Section 3. Precisely, for all $x \in \mathbb{H}$, one can consider the algorithm $y_{k+1} = \overline{\mathcal{FB}}_{f,g}(x, y_k)$, denoted by (\mathcal{A}_2) , in order to compute numerically $\text{prox}_{f+g}(x)$, with the only knowledge of prox_f and ∇g . Convergence proof (under some assumptions on f and g) of (\mathcal{A}_2) should be the topic of a future work.

In Section 4.2 we turn back to our initial motivation, namely the sensitivity analysis of the parameterized variational inequality (1). Precisely, under some assumptions (see Proposition 4.3 for details), we derive from the decomposition formula (3) that if

$$u(t) := \text{prox}_{f+g}(r(t)),$$

for all $t \geq 0$, where $f := \iota_K$ (where $K \subset \mathbb{H}$ is a nonempty closed convex subset) and where $g \in \Gamma_0(\mathbb{H})$ and $r: \mathbb{R}^+ \rightarrow \mathbb{H}$ are smooth enough, then

$$u'(0) = \text{prox}_{\varphi_f + \psi_g}(r'(0)),$$

where $\varphi_f := \iota_C$ (where C is a nonempty closed convex subset of \mathbb{H} related to K) and where $\psi_g(x) := \frac{1}{2} \langle \text{D}^2 g(u(0))(x), x \rangle$ for all $x \in \mathbb{H}$. It should be noted that

the assumptions of Proposition 4.3 are quite restrictive, raising open questions about their relaxations (see Remark 4.5). This also should be the subject of a forthcoming work.

1.2. Notations and basics

In this section we introduce some notations available throughout the paper and we recall some basics of convex analysis. We refer to standard books like [2, 9, 14] and references therein.

Let H be a real Hilbert space and let $\langle \cdot, \cdot \rangle$ (resp. $\|\cdot\|$) be the corresponding scalar product (resp. norm). For every subset S of H , we denote respectively by $\text{int}(S)$ and $\text{cl}(S)$ its *interior* and its *closure*. In the sequel we denote by $I: H \rightarrow H$ the *identity operator* and by $L_x: H \rightarrow H$ the *affine operator* defined by

$$L_x(y) := x - y,$$

for all $x, y \in H$. For a set-valued map $A: H \rightrightarrows H$, the *domain* of A is given by

$$D(A) := \{x \in H \mid A(x) \neq \emptyset\}.$$

We denote by $A^{-1}: H \rightrightarrows H$ the set-valued map defined by

$$A^{-1}(y) := \{x \in H \mid y \in A(x)\},$$

for all $y \in H$. Note that $y \in A(x)$ if and only if $x \in A^{-1}(y)$, for all $x, y \in H$. The *range* of A is given by

$$R(A) := \{y \in H \mid A^{-1}(y) \neq \emptyset\} = D(A^{-1}).$$

We denote by $\text{Fix}(A)$ the set of all fixed points of A , that is, the set given by

$$\text{Fix}(A) := \{x \in H \mid x \in A(x)\}.$$

Finally, if $A(x)$ is a singleton for all $x \in D(A)$, we say that A is *single-valued*.

For all extended-real-valued functions $g: H \rightarrow \mathbb{R} \cup \{+\infty\}$, the *domain* of g is given by

$$\text{dom}(g) := \{x \in H \mid g(x) < +\infty\}.$$

Recall that g is said to be *proper* if $\text{dom}(g) \neq \emptyset$.

Let $g: H \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper extended-real-valued function. We denote by $g^*: H \rightarrow \mathbb{R} \cup \{+\infty\}$ the *conjugate* of g defined by

$$g^*(y) := \sup_{z \in H} \{\langle y, z \rangle - g(z)\},$$

for all $y \in H$. Clearly g^* is lower semicontinuous and convex.

We denote by $\Gamma_0(H)$ the set of all extended-real-valued functions $g: H \rightarrow \mathbb{R} \cup \{+\infty\}$ that are proper, lower semicontinuous and convex. If $g \in \Gamma_0(H)$, we

recall that $g^* \in \Gamma_0(\mathbb{H})$ and that the *Fenchel-Moreau equality* $g^{**} = g$ holds. For all $g \in \Gamma_0(\mathbb{H})$, we denote by $\partial g: \mathbb{H} \rightrightarrows \mathbb{H}$ the *Fenchel-Moreau subdifferential* of g defined by

$$\partial g(x) := \{y \in \mathbb{H} \mid \langle y, z - x \rangle \leq g(z) - g(x), \forall z \in \mathbb{H}\},$$

for all $x \in \mathbb{H}$. It is easy to check that ∂g is a monotone operator and that, for all $x \in \mathbb{H}$, $0 \in \partial g(x)$ if and only if $x \in \operatorname{argmin} g$. Moreover, for all $x, y \in \mathbb{H}$, it holds that $y \in \partial g(x)$ if and only if $x \in \partial g^*(y)$. Recall that, if g is differentiable on \mathbb{H} , then $\partial g(x) = \{\nabla g(x)\}$ for all $x \in \mathbb{H}$.

Let $A: \mathbb{H} \rightarrow \mathbb{H}$ be a single-valued operator defined everywhere on \mathbb{H} , and let $g \in \Gamma_0(\mathbb{H})$. We denote by $\operatorname{VI}(A, g)$ the variational inequality which consists of finding $y \in \mathbb{H}$ such that

$$-A(y) \in \partial g(y), \quad \text{or equivalently, } \langle A(y), z - y \rangle + g(z) - g(y) \geq 0,$$

for all $z \in \mathbb{H}$. Then we denote by $\operatorname{Sol}_{\operatorname{VI}}(A, g)$ the set of solutions of $\operatorname{VI}(A, g)$. Recall that if A is Lipschitzian and strongly monotone, then $\operatorname{VI}(A, g)$ admits a unique solution, *i.e.* $\operatorname{Sol}_{\operatorname{VI}}(A, g)$ is a singleton.

Let $g \in \Gamma_0(\mathbb{H})$. The classical *proximal operator* of g is defined by

$$\operatorname{prox}_g := (\operatorname{I} + \partial g)^{-1}.$$

Recall that prox_g is a single-valued operator defined everywhere on \mathbb{H} . Moreover, it can be characterized as follows:

$$\operatorname{prox}_g(x) = \operatorname{argmin} \left(g + \frac{1}{2} \|\cdot - x\|^2 \right) = \operatorname{Sol}_{\operatorname{VI}}(-L_x, g),$$

for all $x \in \mathbb{H}$. It is also well-known that

$$\operatorname{Fix}(\operatorname{prox}_g) = \operatorname{argmin} g.$$

The classical *Moreau's envelope* $M_g: \mathbb{H} \rightarrow \mathbb{R}$ of g is defined by

$$M_g(x) := \min \left(g + \frac{1}{2} \|\cdot - x\|^2 \right),$$

for all $x \in \mathbb{H}$. Recall that M_g is convex and differentiable on \mathbb{H} with $\nabla M_g = \operatorname{prox}_{g^*}$. Let us also recall the classical Moreau's decompositions

$$\operatorname{prox}_g + \operatorname{prox}_{g^*} = \operatorname{I} \quad \text{and} \quad M_g + M_{g^*} = \frac{1}{2} \|\cdot\|^2.$$

Finally, it is well-known that if $g = \iota_K$ is the *indicator function* of a nonempty closed and convex subset K of \mathbb{H} , that is, $\iota_K(x) = 0$ if $x \in K$ and $\iota_K(x) = +\infty$ if not, then $\operatorname{prox}_g = \operatorname{proj}_K$, where proj_K denotes the classical projection operator on the subset K .

2. The f -proximal operator

2.1. Definition and main result

Let $f, g \in \Gamma_0(\mathbb{H})$. In this section we introduce (see Definition 2.1) a new operator denoted by prox_g^f , generalizing the classical proximal operator prox_g . Assuming that $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$, and under the additivity condition $\partial(f+g) = \partial f + \partial g$, we prove in Theorem 2.8 that prox_{f+g} can be written as the composition of prox_f with prox_g^f .

Definition 2.1. (f -proximal operator) Let $f, g \in \Gamma_0(\mathbb{H})$. The f -proximal operator of g is the set-valued map $\text{prox}_g^f: \mathbb{H} \rightrightarrows \mathbb{H}$ defined by

$$\text{prox}_g^f := (\text{I} + \partial g \circ \text{prox}_f)^{-1}. \quad (5)$$

Note that prox_g^f can be seen as a generalization of prox_g since $\text{prox}_g^c = \text{prox}_g$ for all constant $c \in \mathbb{R}$.

Example 2.2. Let us assume that $\mathbb{H} = \mathbb{R}$. We consider $f = \iota_{[-1,1]}$ and $g(x) = |x|$ for all $x \in \mathbb{R}$. In that case we obtain that $\partial g \circ \text{prox}_f = \partial g$ and thus $\text{prox}_g^f = \text{prox}_g$.

Example 2.2 provides a simple situation where $\text{prox}_g^f = \text{prox}_g$ while f is not constant. We provide in Propositions 2.13 and 2.15 some general sufficient (and necessary) conditions under which $\text{prox}_g^f = \text{prox}_g$.

Example 2.3. Let us assume that $\mathbb{H} = \mathbb{R}$. We consider $f = \iota_{\{0\}}$ and $g(x) = |x|$ for all $x \in \mathbb{R}$. In that case we obtain that $\partial g \circ \text{prox}_f(x) = [-1, 1]$ for all $x \in \mathbb{R}$. As a consequence $\text{prox}_g^f(x) = [x-1, x+1]$ for all $x \in \mathbb{R}$. See Figure 2.1 for graphical representations of prox_g and prox_g^f in that case.

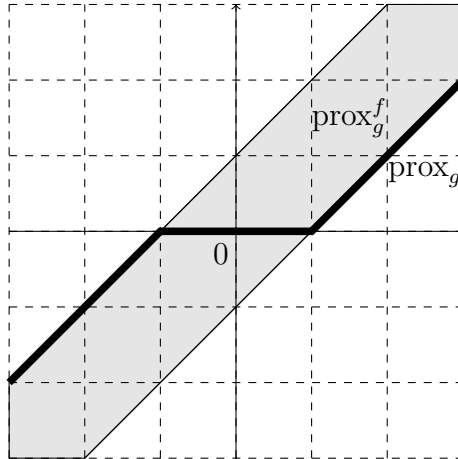


Figure 2.1: Example 2.3, graph of prox_g in bold line, and graph of prox_g^f in gray.

Example 2.4. Let us assume that $\mathbb{H} = \mathbb{R}$. We consider $f(x) = g(x) = |x|$ for all $x \in \mathbb{R}$. In that case we obtain that $\partial g \circ \text{prox}_f(x) = -1$ for all $x < -1$, $\partial g \circ \text{prox}_f(x) = [-1, 1]$ for all $x \in [-1, 1]$ and $\partial g \circ \text{prox}_f(x) = 1$ for all $x > 1$. As a consequence $\text{prox}_g^f(x) = x + 1$ for all $x \leq -2$, $\text{prox}_g^f(x) = [-1, x + 1]$ for all

$x \in [-2, 0]$, $\text{prox}_g^f(x) = [x - 1, 1]$ for all $x \in [0, 2]$ and $\text{prox}_g^f(x) = x - 1$ for all $x \geq 2$. See Figure 2.2 for graphical representations of prox_g and prox_g^f in that case.

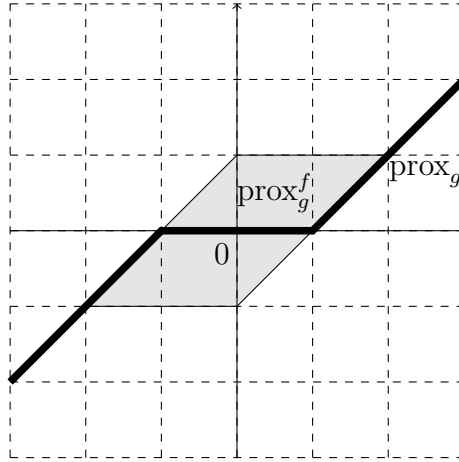


Figure 2.2: Example 2.4, graph of prox_g in bold line, and graph of prox_g^f in gray.

Examples 2.3 and 2.4 provide simple illustrations where prox_g^f is not single-valued. In particular it follows that prox_g^f cannot be written as a proximal operator prox_φ for some $\varphi \in \Gamma_0(H)$. We provide in Proposition 2.18 some sufficient conditions under which prox_g^f is single-valued. Moreover, Examples 2.3 and 2.4 provide simple situations where $\partial g \circ \text{prox}_f$ is not a monotone operator. As a consequence, it may be possible that $D(\text{prox}_g^f) \subsetneq H$. In the next proposition, a necessary and sufficient condition under which $D(\text{prox}_g^f) = H$ is derived.

Proposition 2.5. *Let $f, g \in \Gamma_0(H)$ such that $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$. It holds that $D(\text{prox}_g^f) = H$ if and only if the additivity condition*

$$\partial(f + g) = \partial f + \partial g, \tag{C_1}$$

is satisfied.

Proof. We first assume that $\partial(f + g) = \partial f + \partial g$. Let $x \in H$. Defining $w = \text{prox}_{f+g}(x) \in H$, we obtain that $x \in w + \partial(f + g)(w) = w + \partial f(w) + \partial g(w)$. Thus, there exist $w_f \in \partial f(w)$ and $w_g \in \partial g(w)$ such that $x = w + w_f + w_g$. We define $y = w + w_f \in w + \partial f(w)$. In particular we have $w = \text{prox}_f(y)$. Moreover we obtain $x = y + w_g \in y + \partial g(w) = y + \partial g(\text{prox}_f(y))$. We conclude that $y \in \text{prox}_g^f(x)$.

Without any additional assumption and directly from the definition of the sub-differential, one can easily see that the inclusion $\partial f(w) + \partial g(w) \subset \partial(f + g)(w)$ is always satisfied for every $w \in H$. Now let us assume that $D(\text{prox}_g^f) = H$. Let $w \in H$ and let $z \in \partial(f + g)(w)$. We consider $x = w + z \in w + \partial(f + g)(w)$. In particular it holds that $w = \text{prox}_{f+g}(x)$. Since $D(\text{prox}_g^f) = H$, there exists $y \in \text{prox}_g^f(x)$ and thus it holds that $x \in y + \partial g(\text{prox}_f(y))$. Moreover, since $y \in \text{prox}_f(y) + \partial f(\text{prox}_f(y))$, we get that $x \in \text{prox}_f(y) + \partial f(\text{prox}_f(y)) + \partial g(\text{prox}_f(y)) \subset \text{prox}_f(y) + \partial(f + g)(\text{prox}_f(y))$. Thus it holds that $\text{prox}_f(y) = \text{prox}_{f+g}(x) = w$.

Moreover, since $x \in \text{prox}_f(y) + \partial f(\text{prox}_f(y)) + \partial g(\text{prox}_f(y))$, we obtain that $x \in w + \partial f(w) + \partial g(w)$. We have proved that $z = x - w \in \partial f(w) + \partial g(w)$. This concludes the proof. \square

In most of the present paper, we will assume that Condition (C_1) is satisfied. It is not our aim here to discuss the weakest qualification condition ensuring that condition. A wide literature already deals with this topic (see, e.g., [1, 7, 14]). However, we recall in the following remark the classical sufficient condition of Moreau-Rockafellar under which Condition (C_1) holds true (see, e.g., [2, Corollary 16.48 p.277]), and we provide a simple example where Condition (C_1) does not hold and $D(\text{prox}_g^f) \subsetneq H$.

Remark 2.6. (Moreau-Rockafellar theorem) Let $f, g \in \Gamma_0(H)$ such that we have $\text{dom}(f) \cap \text{int}(\text{dom}(g)) \neq \emptyset$. Then $\partial(f + g) = \partial f + \partial g$.

Example 2.7. Let us assume that $H = \mathbb{R}$. We consider $f = \iota_{\mathbb{R}^-}$ and $g(x) = \iota_{\mathbb{R}^+}(x) - \sqrt{x}$ for all $x \in \mathbb{R}$. In that case, one can easily check that $\text{dom}(f) \cap \text{dom}(g) = \{0\} \neq \emptyset$, $\partial f(0) + \partial g(0) = \emptyset \subsetneq \mathbb{R} = \partial(f + g)(0)$ and $D(\text{prox}_g^f) = \emptyset \subsetneq H$.

We are now in position to state and prove the main result of the present paper.

Theorem 2.8. Let $f, g \in \Gamma_0(H)$ such that $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$. If $\partial(f + g) = \partial f + \partial g$, then the decomposition formula

$$\text{prox}_{f+g} = \text{prox}_f \circ \text{prox}_g^f \quad (6)$$

holds true. In other words, for every $x \in H$, we have $\text{prox}_{f+g}(x) = \text{prox}_f(z)$ for all $z \in \text{prox}_g^f(x)$.

Proof. Let $x \in H$ and let $y \in \text{prox}_g^f(x)$ constructed as in the first part of the proof of Proposition 2.5. In particular it holds that $\text{prox}_f(y) = \text{prox}_{f+g}(x)$. Let $z \in \text{prox}_g^f(x)$. We know that $x - y \in \partial g(\text{prox}_f(y))$ and $x - z \in \partial g(\text{prox}_f(z))$. Since ∂g is a monotone operator, we obtain that

$$\langle (x - y) - (x - z), \text{prox}_f(y) - \text{prox}_f(z) \rangle \geq 0.$$

From the cocoercivity (see for instance [2, Definition 4.10 p.72]) of the proximal operator, we obtain that

$$0 \geq \langle y - z, \text{prox}_f(y) - \text{prox}_f(z) \rangle \geq \|\text{prox}_f(y) - \text{prox}_f(z)\|^2 \geq 0.$$

We deduce that $\text{prox}_f(z) = \text{prox}_f(y) = \text{prox}_{f+g}(x)$. The proof is complete. \square

Remark 2.9. Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and such that $\partial(f + g) = \partial f + \partial g$ and let $x \in H$. Theorem 2.8 states that, even if $\text{prox}_g^f(x)$ is not a singleton, all elements of $\text{prox}_g^f(x)$ has the same value through the proximal operator prox_f , and this value is equal to $\text{prox}_{f+g}(x)$.

Remark 2.10. Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$. Note that the additivity condition $\partial(f + g) = \partial f + \partial g$ is not only sufficient, but also necessary for

the validity of the equality $\text{prox}_{f+g} = \text{prox}_f \circ \text{prox}_g^f$. Indeed, from Proposition 2.5, if $\partial f + \partial g \subsetneq \partial(f+g)$, then there exists $x \in H$ such that $\text{prox}_g^f(x) = \emptyset$ and thus $\text{prox}_{f+g}(x) \neq \text{prox}_f \circ \text{prox}_g^f(x)$.

Remark 2.11. Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and such that $\partial(f+g) = \partial f + \partial g$. From Theorem 2.8, we deduce that $R(\text{prox}_{f+g}) \subset R(\text{prox}_f) \cap R(\text{prox}_g)$. If the additivity condition $\partial(f+g) = \partial f + \partial g$ is not satisfied, this remark does not hold true anymore. Indeed, with the framework of Example 2.7, we have $R(\text{prox}_{f+g}) = \{0\}$ while $0 \notin R(\text{prox}_g)$.

Example 2.12. Following the idea of Yu in [18, Example 2], let us consider $H = \mathbb{R}$ and $f(x) = \frac{1}{2}x^2$ for all $x \in \mathbb{R}$. Since $\text{prox}_{\gamma f} = \frac{1}{1+\gamma}I$ for all $\gamma \geq 0$ and $\text{prox}_f^f = \frac{2}{3}I$, we retrieve that

$$\frac{1}{3}I = \text{prox}_{2f} = \text{prox}_{f+f} = \text{prox}_f \circ \text{prox}_f^f = \frac{1}{3}I \neq \frac{1}{4}I = \text{prox}_f \circ \text{prox}_f,$$

which illustrates Theorem 2.8.

2.2. Properties

Let $f, g \in \Gamma_0(H)$. We know that prox_g^f is a generalization of prox_g in the sense that $\text{prox}_g^f = \text{prox}_g$ if f is constant for instance. In the next proposition, our aim is to provide more general sufficient (and necessary) conditions under which $\text{prox}_g^f = \text{prox}_g$. We will base our discussion on the following conditions:

$$\forall x \in H, \quad \partial g(x) \subset \partial g(\text{prox}_f(x)), \tag{C_2}$$

$$\forall x \in H, \quad \partial g(\text{prox}_f(x)) \subset \partial g(x). \tag{C_3}$$

Note that Condition (C₂) has been introduced by Yu in [18] as a sufficient condition under which $\text{prox}_{f+g} = \text{prox}_f \circ \text{prox}_g$.

Proposition 2.13. *Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$.*

- (i) *If Condition (C₂) is satisfied, then $\text{prox}_g(x) \in \text{prox}_g^f(x)$ for all $x \in H$.*
- (ii) *If Conditions (C₁) and (C₃) are satisfied, then $\text{prox}_g^f(x) = \text{prox}_g(x)$ for all $x \in H$.*

In both cases, Condition (C₁) is satisfied and the equality $\text{prox}_{f+g} = \text{prox}_f \circ \text{prox}_g$ holds true.

Proof. Let $x \in H$. If Condition (C₂) is satisfied, considering $y = \text{prox}_g(x)$, we get that $x \in y + \partial g(y) \subset y + \partial g(\text{prox}_f(y))$ and thus $y \in \text{prox}_g^f(x)$. In particular, it holds that $D(\text{prox}_g^f) = H$ and thus Condition (C₁) is satisfied from Proposition 2.5. Secondly, if Conditions (C₁) and (C₃) are satisfied, then $D(\text{prox}_g^f) = H$ from Proposition 2.5. Considering $y \in \text{prox}_g^f(x)$, we get that $x \in y + \partial g(\text{prox}_f(y)) \subset y + \partial g(y)$ and thus $y = \text{prox}_g(x)$. The last assertion of Proposition 2.13 directly follows from Theorem 2.8. □

In the first item of Proposition 2.13 and if prox_g^f is set-valued, we are in the situation where prox_g is a selection of prox_g^f . Proposition 2.15 specifies this selection in the case where $\partial(f+g) = \partial f + \partial g$.

Lemma 2.14. *Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$. Then $\text{prox}_g^f(x)$ is a nonempty closed and convex subset of H for all $x \in D(\text{prox}_g^f)$.*

Proof. The proof of Lemma 2.14 is provided after the proof of Proposition 3.2 (required). \square

Proposition 2.15. *Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and such that $\partial(f+g) = \partial f + \partial g$ and let $x \in H$. If $\text{prox}_g(x) \in \text{prox}_g^f(x)$, then*

$$\text{prox}_g(x) = \text{proj}_{\text{prox}_g^f(x)}(\text{prox}_{f+g}(x)).$$

Proof. If $\text{prox}_g(x) \in \text{prox}_g^f(x)$, then $x \in D(\text{prox}_g^f)$ and thus $\text{prox}_g^f(x)$ is a nonempty closed and convex subset of H from Lemma 2.14. Let $z \in \text{prox}_g^f(x)$. In particular we have $\text{prox}_f(z) = \text{prox}_{f+g}(x)$ from Theorem 2.8. Using the fact that $x - \text{prox}_g(x) \in \partial g(\text{prox}_g(x))$ and $x - z \in \partial g(\text{prox}_f(z)) = \partial g(\text{prox}_{f+g}(x))$ together with the monotonicity of ∂g , we obtain that $\langle \text{prox}_{f+g}(x) - \text{prox}_g(x), z - \text{prox}_g(x) \rangle = \langle \text{prox}_{f+g}(x) - \text{prox}_g(x), (x - \text{prox}_g(x)) - (x - z) \rangle \leq 0$. Since $\text{prox}_g(x) \in \text{prox}_g^f(x)$, we conclude the proof from the characterization of $\text{proj}_{\text{prox}_g^f(x)}$. \square

Remark 2.16. Let $f = \iota_{\{\omega\}}$ with $\omega \in H$ and let $g \in \Gamma_0(H)$ such that $\omega \in \text{int}(\text{dom}(g))$. Hence the additivity condition $\partial(f+g) = \partial f + \partial g$ is satisfied from Remark 2.6. From Remark 2.11 and since $\text{prox}_f = \text{proj}_{\{\omega\}}$, we easily deduce that $R(\text{prox}_{f+g}) = \{\omega\}$. Let $x \in H$ such that $\text{prox}_g(x) \in \text{prox}_g^f(x)$. From Proposition 2.15 we get that

$$\text{prox}_g(x) = \text{proj}_{\text{prox}_g^f(x)}(\omega).$$

If moreover $\omega = 0$, we deduce that $\text{prox}_g(x)$ is the particular selection that corresponds to the element of minimal norm in $\text{prox}_g^f(x)$ (also known as the *lazy selection*). The following example is in this sense.

Example 2.17. Let us consider the framework of Example 2.3. In that case, Conditions (C₁) and (C₂) are satisfied. We deduce from Proposition 2.13 that $\text{prox}_g(x) \in \text{prox}_g^f(x)$ for all $x \in \mathbb{R}$. From Remark 2.16, we conclude that $\text{prox}_g(x)$ is exactly the element of minimal norm in $\text{prox}_g^f(x)$ for all $x \in \mathbb{R}$. This result is clearly illustrated by the graphs of prox_g and prox_g^f provided in Figure 2.1.

Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and such that $\partial(f+g) = \partial f + \partial g$. From Theorem 2.8, one can easily see that, if prox_f is injective, then prox_g^f is single-valued. Since the injection of prox_f is too restrictive, other sufficient conditions under which prox_g^f is single-valued are provided from Theorem 2.8 in the next proposition.

Proposition 2.18. *Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and such that $\partial(f+g) = \partial f + \partial g$. If either ∂f or ∂g is single-valued, then prox_g^f is single-valued.*

Proof. Let $x \in H$ and let $z_1, z_2 \in \text{prox}_g^f(x)$. From Theorem 2.8, it holds that $\text{prox}_f(z_1) = \text{prox}_f(z_2) = \text{prox}_{f+g}(x)$. If the operator ∂f is single-valued, we obtain that $z_1 = \text{prox}_{f+g}(x) + \partial f(\text{prox}_{f+g}(x)) = z_2$. If the operator ∂g is single-valued, we get $x - z_1 = \partial g(\text{prox}_f(z_1)) = \partial g(\text{prox}_f(z_2)) = x - z_2$ and thus $z_1 = z_2$. \square

3. Relationship with the Douglas-Rachford operator

Let $f, g \in \Gamma_0(\mathbb{H})$. The *Douglas-Rachford operator* $\mathcal{DR}_{f,g}: \mathbb{H} \rightarrow \mathbb{H}$ associated to f and g is usually defined by

$$\mathcal{DR}_{f,g}(y) := y - \text{prox}_f(y) + \text{prox}_g(2\text{prox}_f(y) - y),$$

for all $y \in \mathbb{H}$. We refer for instance to [2, Section 28.3 p.517] where details can be found on this classical operator.

One aim of this section is to study the relations between the f -proximal operator prox_g^f introduced in this paper and the Douglas-Rachford operator $\mathcal{DR}_{f,g}$. For this purpose, we introduce an extension $\overline{\mathcal{DR}}_{f,g}: \mathbb{H} \times \mathbb{H} \rightarrow \mathbb{H}$ of the classical Douglas-Rachford operator defined by

$$\overline{\mathcal{DR}}_{f,g}(x, y) := y - \text{prox}_f(y) + \text{prox}_g(x + \text{prox}_f(y) - y),$$

for all $x, y \in \mathbb{H}$. Note that $\mathcal{DR}_{f,g}(y) = \overline{\mathcal{DR}}_{f,g}(\text{prox}_f(y), y)$ for all $y \in \mathbb{H}$, and that the definition of $\overline{\mathcal{DR}}_{f,g}$ only depends on the knowledge of prox_f and prox_g .

3.1. Several characterizations of prox_g^f

Let $f, g \in \Gamma_0(\mathbb{H})$. In this subsection, our aim is to derive several characterizations of prox_g^f in terms of solutions of variational inequalities, of minimization problems and of fixed point problems (see Proposition 3.2).

Lemma 3.1. *Let $f, g \in \Gamma_0(\mathbb{H})$. For all $x \in \mathbb{H}$ it holds that*

$$\overline{\mathcal{DR}}_{f,g}(x, \cdot) = \text{prox}_{g^* \circ L_x} \circ \text{prox}_{f^*}.$$

Proof. Let $x \in \mathbb{H}$. Lemma 3.1 directly follows from the equality $\text{prox}_{g^* \circ L_x} = L_x \circ \text{prox}_{g^*} \circ L_x$ (see [2, Proposition 24.8 p.416]) and from Moreau's decompositions. \square

Proposition 3.2. *Let $f, g \in \Gamma_0(\mathbb{H})$. For all $x \in \mathbb{H}$ it holds that*

$$\text{prox}_g^f(x) = \text{Sol}_{\text{VI}}(\text{prox}_f, g^* \circ L_x) = \text{argmin}(\text{M}_{f^*} + g^* \circ L_x) = \text{Fix}(\overline{\mathcal{DR}}_{f,g}(x, \cdot)).$$

Proof. In this proof we will use standard properties of convex analysis recalled in Section 1.2. Let $x \in \mathbb{H}$. One can easily prove that $\partial(g^* \circ L_x) = -\partial g^* \circ L_x$. For all $y \in \mathbb{H}$, it holds that

$$\begin{aligned} y \in \text{prox}_g^f(x) &\iff x - y \in \partial g(\text{prox}_f(y)) \iff \text{prox}_f(y) \in \partial g^*(x - y) \\ &\iff -\text{prox}_f(y) \in \partial(g^* \circ L_x)(y). \end{aligned}$$

Moreover, since $\text{dom}(\text{M}_{f^*}) = \mathbb{H}$ and from Remark 2.6, we have

$$\begin{aligned} -\text{prox}_f(y) \in \partial(g^* \circ L_x)(y) &\iff 0 \in \nabla \text{M}_{f^*}(y) + \partial(g^* \circ L_x)(y) \\ &\iff 0 \in \partial(\text{M}_{f^*} + g^* \circ L_x)(y). \end{aligned}$$

$$\begin{aligned} \text{Finally: } -\text{prox}_f(y) \in \partial(g^* \circ L_x)(y) &\iff \text{prox}_{f^*}(y) \in y + \partial(g^* \circ L_x)(y) \\ &\iff y = \text{prox}_{g^* \circ L_x} \circ \text{prox}_{f^*}(y). \end{aligned}$$

This concludes the proof from Lemma 3.1. \square

Proof of Lemma 2.14. Let $x \in D(\text{prox}_g^f)$. In particular $\text{prox}_g^f(x)$ is not empty. From Proposition 3.2, we have

$$\text{prox}_g^f(x) = \text{argmin}(M_{f^*} + g^* \circ L_x).$$

Since $M_{f^*} + g^* \circ L_x \in \Gamma_0(\mathbb{H})$, one can easily deduce that $\text{prox}_g^f(x)$ is closed and convex. \square

3.2. A weakly convergent algorithm that computes prox_g^f numerically

Let $f, g \in \Gamma_0(\mathbb{H})$. In this section, our aim is to derive from Proposition 3.2 an algorithm, that depends only on the knowledge of prox_f and prox_g , allowing to compute numerically an element of $\text{prox}_g^f(x)$ for all $x \in D(\text{prox}_g^f)$. We refer to Algorithm (\mathcal{A}_1) in Theorem 3.3.

Moreover, if the additivity condition $\partial(f + g) = \partial f + \partial g$ is satisfied, it follows from Theorem 2.8 that Algorithm (\mathcal{A}_1) is an algorithm allowing to compute numerically $\text{prox}_{f+g}(x)$ for all $x \in \mathbb{H}$ with the only knowledge of prox_f and prox_g .

Theorem 3.3. *Let $f, g \in \Gamma_0(\mathbb{H})$ and let $x \in D(\text{prox}_g^f)$ be fixed. Then, Algorithm (\mathcal{A}_1) given by*

$$y_0 \in \mathbb{H}, \quad y_{k+1} = \overline{\mathcal{DR}}_{f,g}(x, y_k), \quad (\mathcal{A}_1)$$

weakly converges to an element $y^ \in \text{prox}_g^f(x)$. Moreover, if $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and $\partial(f + g) = \partial f + \partial g$, it holds that $\text{prox}_f(y^*) = \text{prox}_{f+g}(x)$.*

Proof. From Lemma 3.1, $\overline{\mathcal{DR}}_{f,g}(x, \cdot)$ coincides with the composition of two firmly non-expansive operators, and thus of two non-expansive and $\frac{1}{2}$ -averaged operators (see [2, Remark 4.34(iii) p.81]). Since $x \in D(\text{prox}_g^f)$, it follows from Proposition 3.2 and Lemma 3.1 that $\text{Fix}(\text{prox}_{g^* \circ L_x} \circ \text{prox}_{f^*}) \neq \emptyset$. We conclude from [2, Theorem 5.23 p.100] that Algorithm (\mathcal{A}_1) weakly converges to a fixed point y^* of $\overline{\mathcal{DR}}_{f,g}(x, \cdot)$. From Proposition 3.2, it holds that $y^* \in \text{prox}_g^f(x)$. Finally, if $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and $\partial(f + g) = \partial f + \partial g$, we conclude that $\text{prox}_f(y^*) = \text{prox}_{f+g}(x)$ from Theorem 2.8. \square

Remark 3.4. As already mentioned in the introduction, it turns out that Algorithm (\mathcal{A}_1) was already considered, up to some translations, and implemented in previous works (see, e.g., the so-called *dual forward-backward splitting* in [4, Algorithm 3.5]), showing that the f -proximal operator prox_g^f is already present (in a hidden form) and useful for numerical purposes in the existing literature. However, to the best of our knowledge, it has never been explicitly expressed in a closed formula such as (5) and neither been deeply studied from a theoretical point of view.

Remark 3.5. Let us discuss with more details the relationship between the present work and the one proposed in [4]. Let $x \in \mathbb{H}$. In [4, Proposition 3.4], the authors prove that if $v \in \mathbb{H}$ is a solution to

$$\text{argmin}(M_{f^*} \circ L_x + g^*) \quad (7)$$

then $\text{prox}_f(x - v) = \text{prox}_{f+g}(x)$. Combining this result with Proposition 3.2 easily constitutes an alternative proof of the new decomposition formula (6) derived in this paper. Moreover, in [4, Algorithm 3.5], the authors consider the so-called *dual forward-backward splitting* given by

$$v_{k+1} = \text{prox}_{g^*}(v_k + \text{prox}_f(x - v_k)),$$

which is related to Algorithm (\mathcal{A}_1) by setting $y_k = x - v_k$. From Proposition 3.2, the present work points out that the operator given in (7) actually coincides, up to a translation, with a generalization of the classical proximal operator, that is exactly the f -proximal operator introduced and studied from a theoretical point of view in this paper. In this section we also prove that Algorithm (\mathcal{A}_1) actually coincides with a fixed-point algorithm associated to a generalized version of the classical Douglas-Rachford operator.

Remark 3.6. Let $f, g \in \Gamma_0(\mathbb{H})$ and let $x \in \text{D}(\text{prox}_g^f)$. Algorithm (\mathcal{A}_1) consists in a fixed-point algorithm from the characterization given in Proposition 3.2 by

$$\text{prox}_g^f(x) = \text{Fix}(\overline{\mathcal{DR}}_{f,g}(x, \cdot)).$$

Actually, one can easily see that Algorithm (\mathcal{A}_1) also coincides with the well-known *Forward-Backward algorithm* (see [5, Section 10.3 p.191] for details) from the characterization given in Proposition 3.2 by

$$\text{prox}_g^f(x) = \text{argmin}(M_{f^*} + g^* \circ L_x).$$

Indeed, we recall that M_{f^*} is differentiable with $\nabla M_{f^*} = \text{prox}_f$. We also refer to Section 4.1 for a brief discussion about the Forward-Backward algorithm.

3.3. Recovering a classical result from the decomposition formula

Let $f, g \in \Gamma_0(\mathbb{H})$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and such that $\partial(f+g) = \partial f + \partial g$. Our aim in this section is to recover in a simple way, with the help of the decomposition formula (6), the well-known equality

$$\text{argmin } f + g = \text{prox}_f(\text{Fix}(\mathcal{DR}_{f,g})).$$

This result can be found for example in [2, Proposition 26.1].

Lemma 3.7. *Let $f, g \in \Gamma_0(\mathbb{H})$. It holds that*

$$\text{Fix}(\mathcal{DR}_{f,g}) = \text{Fix}(\text{prox}_g^f \circ \text{prox}_f).$$

Proof. Let $z \in \mathbb{H}$. It holds from Proposition 3.2 that

$$\begin{aligned} z \in \text{Fix}(\mathcal{DR}_{f,g}) &\iff z = \mathcal{DR}_{f,g}(z) = \overline{\mathcal{DR}}_{f,g}(\text{prox}_f(z), z) \\ &\iff z \in \text{Fix}(\overline{\mathcal{DR}}_{f,g}(\text{prox}_f(z), \cdot)) = \text{prox}_g^f(\text{prox}_f(z)) \\ &\iff z \in \text{Fix}(\text{prox}_g^f \circ \text{prox}_f). \quad \square \end{aligned}$$

Proposition 3.8. *Let $f, g \in \Gamma_0(H)$ with $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ and such that $\partial(f+g) = \partial f + \partial g$. It holds that*

$$\text{argmin } f + g = \text{prox}_f(\text{Fix}(\mathcal{DR}_{f,g})).$$

Proof. Let $y \in \text{Fix}(\mathcal{DR}_{f,g})$. Then $y \in \text{Fix}(\text{prox}_g^f \circ \text{prox}_f)$ from Lemma 3.7. Thus $y \in \text{prox}_g^f \circ \text{prox}_f(y)$. From the decomposition formula (6), we get that $\text{prox}_f(y) = \text{prox}_{f+g}(\text{prox}_f(y))$ and thus $\text{prox}_f(y) \in \text{argmin } f + g$.

Let $x \in \text{argmin } f + g$. Since $D(\text{prox}_g^f) = H$ from Proposition 2.5, we consider $y \in \text{prox}_g^f(x)$. From the decomposition formula (6) we have $x = \text{prox}_{f+g}(x) = \text{prox}_f(y)$. Let us prove that $y \in \text{Fix}(\mathcal{DR}_{f,g})$. Since $y \in \text{prox}_g^f(x) = \text{Fix}(\overline{\mathcal{DR}}_{f,g}(x, \cdot))$, we get $y = \overline{\mathcal{DR}}_{f,g}(x, y) = \overline{\mathcal{DR}}_{f,g}(\text{prox}_f(y), y) = \mathcal{DR}_{f,g}(y)$. \square

4. Some other applications and forthcoming works

This section can be seen as a conclusion of the paper. Its aim is to provide a glimpse of some other applications of our main result (Theorem 2.8) and to raise open questions for forthcoming works. This section is splitted into two parts.

4.1. Relationship with the Forward-Backward operator

Let $f, g \in \Gamma_0(H)$ such that g is differentiable on H . In that situation, note that the additivity condition $\partial(f+g) = \partial f + \partial g$ is satisfied from Remark 2.6, and that Proposition 2.18 implies that prox_g^f is single-valued.

In that framework, the classical *Forward-Backward operator* $\mathcal{FB}_{f,g}: H \rightarrow H$ associated to f and g is usually defined by

$$\mathcal{FB}_{f,g}(y) := \text{prox}_f(y - \nabla g(y)),$$

for all $y \in H$. We refer to [5, Section 10.3 p.191] for more details. Let us introduce the extension $\overline{\mathcal{FB}}_{f,g}: H \times H \rightarrow H$ defined by

$$\overline{\mathcal{FB}}_{f,g}(x, y) := \text{prox}_f(x - \nabla g(y)),$$

for all $x, y \in H$. In particular, it holds that $\mathcal{FB}_{f,g}(y) = \overline{\mathcal{FB}}_{f,g}(y, y)$ for all $y \in H$. The following result follows from the decomposition formula (6) in Theorem 2.8.

Proposition 4.1. *Let $f, g \in \Gamma_0(H)$ such that g is differentiable on H . Then,*

$$\text{prox}_{f+g}(x) = \text{Fix}(\overline{\mathcal{FB}}_{f,g}(x, \cdot)) \quad \text{for all } x \in H.$$

Proof. Let $x \in H$. Firstly, let $z = \text{prox}_{f+g}(x)$ and let $y = \text{prox}_g^f(x)$. In particular, we have $x = y + \nabla g(\text{prox}_f(y))$. From the decomposition formula (6), we get that $z = \text{prox}_f(y) = \text{prox}_f(x - \nabla g(\text{prox}_f(y))) = \text{prox}_f(x - \nabla g(z)) = \overline{\mathcal{FB}}_{f,g}(x, z)$. Conversely, let $z \in \text{Fix}(\overline{\mathcal{FB}}_{f,g}(x, \cdot))$, that is, $z = \text{prox}_f(x - \nabla g(z))$. Considering $y = x - \nabla g(z)$, we have $z = \text{prox}_f(y)$ and thus $x = y + \nabla g(\text{prox}_f(y))$, that is, $y = \text{prox}_g^f(x)$. Finally, from the decomposition formula (6), we get that $z = \text{prox}_f \circ \text{prox}_g^f(x) = \text{prox}_{f+g}(x)$. \square

From Proposition 4.1, we retrieve the following classical result (see, e.g., [2, Proposition 26.1]).

Proposition 4.2. *Let $f, g \in \Gamma_0(\mathbb{H})$ such that g is differentiable on \mathbb{H} . Then*

$$\operatorname{argmin} f + g = \operatorname{Fix}(\mathcal{FB}_{f,g}).$$

Proof. Let $x \in \mathbb{H}$. It holds that

$$\begin{aligned} x \in \operatorname{argmin} f + g &\iff x = \operatorname{prox}_{f+g}(x) \iff x \in \operatorname{Fix}(\overline{\mathcal{FB}}_{f,g}(x, \cdot)) \\ &\iff x = \overline{\mathcal{FB}}_{f,g}(x, x) = \mathcal{FB}_{f,g}(x) \iff x \in \operatorname{Fix}(\mathcal{FB}_{f,g}). \quad \square \end{aligned}$$

Take $f, g \in \Gamma_0(\mathbb{H})$ such that g is differentiable on \mathbb{H} . The classical *Forward-Backward algorithm* $x_{n+1} = \mathcal{FB}_{f,g}(x_n)$ is a powerful tool since it provides an algorithm, only requiring the knowledge of prox_f and ∇g , that weakly converges (under some conditions on g , see [2, Section 28.5 p.522] for details) to a fixed point of $\mathcal{FB}_{f,g}$, and thus to a minimizer of $f + g$.

From Proposition 4.1, and for all $x \in \mathbb{H}$, one can consider the algorithm (potentially weakly convergent) given by

$$y_0 \in \mathbb{H}, \quad y_{k+1} = \overline{\mathcal{FB}}_{f,g}(x, y_k), \tag{A_2}$$

in order to compute numerically $\operatorname{prox}_{f+g}(x)$, with the only knowledge of prox_f and ∇g . Convergence proof (under some assumptions on f and g) of Algorithm (A₂) should be the topic of a future work.

4.2. Application to sensitivity analysis of variational inequalities

As a conclusion of the present paper, we turn back to our initial motivation, namely the sensitivity analysis, with respect to a nonnegative parameter $t \geq 0$, of some parameterized linear variational inequalities of second kind in a real Hilbert space \mathbb{H} . More precisely, for all $t \geq 0$, we consider the variational inequality which consists of finding $u(t) \in \mathbb{K}$ such that

$$\langle u(t), z - u(t) \rangle + g(z) - g(u(t)) \geq \langle r(t), z - u(t) \rangle,$$

for all $z \in \mathbb{K}$, where $\mathbb{K} \subset \mathbb{H}$ is a nonempty closed and convex set of constraints, and where $g \in \Gamma_0(\mathbb{H})$ and $r: \mathbb{R}^+ \rightarrow \mathbb{H}$ are assumed to be given and smooth enough. The above problem admits a unique solution given by

$$u(t) = \operatorname{prox}_{f+g}(r(t)),$$

where $f = \iota_{\mathbb{K}}$ is the indicator function of \mathbb{K} .

Our aim is to provide from Theorem 2.8 a simple and compact formula for the derivative $u'(0)$ under some assumptions (see Proposition 4.3 for details). Following the idea of Mignot in [10] (see also [8, Theorem 2 p. 620]), we first introduce the following sets

$$\begin{aligned} O_v &:= \{w \in \mathbb{H} \mid \exists \lambda > 0, \operatorname{proj}_{\mathbb{K}}(v) + \lambda w \in \mathbb{K}\} \cap [v - \operatorname{proj}_{\mathbb{K}}(v)]^\perp, \\ C_v &:= \operatorname{cl}\left(\{w \in \mathbb{H} \mid \exists \lambda > 0, \operatorname{proj}_{\mathbb{K}}(v) + \lambda w \in \mathbb{K}\}\right) \cap [v - \operatorname{proj}_{\mathbb{K}}(v)]^\perp, \end{aligned}$$

for all $v \in \mathbb{H}$, where \perp denotes the classical orthogonal of a set.

Proposition 4.3. *Let $v(t) := r(t) - \nabla g(u(t))$ for all $t \in \mathbb{R}$. If the following conditions are satisfied:*

- (i) r is differentiable at $t = 0$;
- (ii) g is twice differentiable on \mathbb{H} ;
- (iii) $O_{v(0)}$ is dense in $C_{v(0)}$;
- (iv) u is differentiable at $t = 0$;

then the derivative $u'(0)$ is given by $u'(0) = \text{prox}_{\varphi_f + \psi_g}(r'(0))$, where $\varphi_f := \iota_{C_{v(0)}}$ and $\psi_g(x) := \frac{1}{2} \langle D^2 g(u(0))(x), x \rangle$ for all $x \in \mathbb{H}$.

Proof. Note that v is differentiable at $t = 0$ with

$$v'(0) = r'(0) - D^2 g(u(0))(u'(0)).$$

Note that prox_g^f is single-valued from Proposition 2.18 and Remark 2.6. From the decomposition formula (6) in Theorem 2.8, one can easily obtain

$$v(t) = \text{prox}_g^f(r(t)), \quad \text{and thus} \quad u(t) = \text{prox}_f \circ \text{prox}_g^f(r(t)) = \text{proj}_K(v(t)),$$

for all $t \geq 0$. Since $O_{v(0)}$ is dense in $C_{v(0)}$, we use the asymptotic expansion of Mignot [10, Theorem 2.1 p. 145] and we obtain

$$\begin{aligned} u'(0) &= \text{proj}_{C_{v(0)}}(v'(0)), \quad \text{and consequently} \\ v'(0) + D^2 g(u(0)) \circ \text{proj}_{C_{v(0)}}(v'(0)) &= r'(0). \end{aligned}$$

Since g is convex and since $C_{v(0)}$ is a nonempty closed convex subset of \mathbb{H} , we deduce that $\varphi_f, \psi_g \in \Gamma_0(\mathbb{H})$. Moreover $\partial(\varphi_f + \psi_g) = \partial\varphi_f + \partial\psi_g$ from Remark 2.6 and $\text{prox}_{\psi_g}^{\varphi_f}$ is single-valued from Proposition 2.18. It also should be noted that $\nabla\psi_g = D^2 g(u(0))$. As a consequence, we have obtained that

$$v'(0) + \nabla\psi_g \circ \text{prox}_{\varphi_f}(v'(0)) = r'(0),$$

that is, $v'(0) = \text{prox}_{\psi_g}^{\varphi_f}(r'(0))$. We conclude the proof from the equality $u'(0) = \text{prox}_{\varphi_f}(v'(0))$ and from Theorem 2.8. \square

Remark 4.4. Proposition 4.3 provides an expression of $u'(0)$ in terms of the proximal operator of a sum of two proper, lower semicontinuous and convex functions. Hence, it could be numerically computed from Algorithm (\mathcal{A}_1) , requiring the knowledge of $\text{proj}_{C_{v(0)}}$ and prox_{ψ_g} . Alternatively, if the convergence is proved, one can also consider Algorithm (\mathcal{A}_2) requiring the knowledge of $\text{proj}_{C_{v(0)}}$ and $\nabla\psi_g = D^2 g(u(0))$.

Remark 4.5. The relaxations in special frameworks of the assumptions of Proposition 4.3 should be the subject of future works. In particular, it would be relevant to provide sufficient conditions ensuring that u is differentiable at $t = 0$. A promising idea in this sense is to invoke the concepts of *twice epi-differentiability* and *proto-differentiability* introduced by Rockafellar in [15, 16].

The application of Proposition 4.3 in the context of some shape optimization problems with unilateral contact and friction is the subject of a forthcoming paper.

A. A nonexistence result for a closed formula

The aim of the present appendix is to prove that there is no closed formula, independent of f and g , allowing to write prox_{f+g} as a linear combination of compositions of linear combinations of I , prox_f , prox_g , prox_f^{-1} and prox_g^{-1} .

For this purpose, we introduce the elementary operator $P_{f,g}^\mu : \text{H} \rightrightarrows \text{H}$ defined by

$$P_{f,g}^\mu := a \text{I} + b \text{prox}_f + c \text{prox}_g + d \text{prox}_f^{-1} + e \text{prox}_g^{-1},$$

for all $\mu = (a, b, c, d, e) \in \mathbb{R}^5$, all $f, g \in \Gamma_0(\text{H})$ and all Hilbert spaces H . Let us assume by contradiction that for all Hilbert spaces H , there exist $m, n \in \mathbb{N}^*$, $\lambda = (\lambda_i)_i \in \mathbb{R}^m$ and $(\mu_{ij})_{ij} \in (\mathbb{R}^5)^{m \times n}$ such that

$$\text{prox}_{f+g} = \sum_{i=1}^m \lambda_i \left(\prod_{j=1}^n P_{f,g}^{\mu_{ij}} \right), \quad (8)$$

for all $f, g \in \Gamma_0(\text{H})$, where \prod denotes finite composition of operators.

Then, we consider the one-dimensional setting $\text{H} = \mathbb{R}$ with $f(x) = g(x) = \frac{\gamma}{2}x^2$ for all $x \in \mathbb{R}$ and all $\gamma \geq 0$. In that case prox_{f+g} is the linear function with slope $\frac{1}{1+2\gamma}$ and each $P_{f,g}^{\mu_{ij}}$ is the linear function with slope

$$\frac{(b_{ij} + c_{ij}) + a_{ij}(1 + \gamma) + (d_{ij} + e_{ij})(1 + \gamma)^2}{1 + \gamma},$$

for all $\gamma \geq 0$. We deduce from equality (8) that

$$\frac{1}{1 + 2\gamma} = \sum_{i=1}^m \lambda_i \left(\prod_{j=1}^n \frac{(b_{ij} + c_{ij}) + a_{ij}(1 + \gamma) + (d_{ij} + e_{ij})(1 + \gamma)^2}{1 + \gamma} \right),$$

for all $\gamma \geq 0$, where \prod denotes now the classical finite product of real numbers. We get that

$$(1 + \gamma)^n = (1 + 2\gamma) \sum_{i=1}^m \lambda_i \left(\prod_{j=1}^n (b_{ij} + c_{ij}) + a_{ij}(1 + \gamma) + (d_{ij} + e_{ij})(1 + \gamma)^2 \right),$$

for all $\gamma \geq 0$. We easily deduce that the above polynomial equality can be extended to all $\gamma \in \mathbb{R}$, and thus it raises a contradiction for $\gamma = -\frac{1}{2}$.

References

- [1] H. Attouch, H. Brezis: Duality for the sum of convex functions in general Banach spaces, in: *Aspects of Mathematics and its Applications*, North-Holland Math. Library 34, North-Holland, Amsterdam (1986) 125–133.
- [2] H. H. Bauschke, P. L. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd edition, CMS Books in Mathematics, Springer, New York (2017).
- [3] A. Bensoussan, J.-L. Lions: Inéquations quasi-variationnelles dépendant d'un paramètre, *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 4(2) (1977) 231–255.
- [4] P. L. Combettes, D. Dũng, B. C. Vũ: Dualization of signal recovery problems, *Set-Valued Var. Anal.* 18(3-4) (2010) 373–404.
- [5] P. L. Combettes, J.-C. Pesquet: Proximal splitting methods in signal processing, in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optim. Appl. 49, Springer, New York (2011) 185–212.
- [6] J. Douglas, H. H. Rachford: On the numerical solution of heat conduction problems in two and three space variables, *Trans. Amer. Math. Soc.* 82 (1956) 421–439.
- [7] I. Ekeland, R. Temam: *Analyse Convexe et Problèmes Variationnels*, Collection Études Mathématiques, Dunod-Gauthier-Villars, Paris et al. (1974).
- [8] A. Haraux: How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities, *J. Math. Soc. Japan* 29(4) (1977) 615–631.
- [9] J.-B. Hiriart-Urruty, C. Lemaréchal: *Fundamentals of Convex Analysis*, Grundlehren Text Editions, Springer, Berlin (2001).
- [10] F. Mignot: Contrôle dans les inéquations variationnelles elliptiques, *J. Functional Analysis* 22(2) (1976) 130–185.
- [11] J.-J. Moreau: Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris* 255 (1962) 2897–2899.
- [12] J.-J. Moreau: Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France* 93 (1965) 273–299.
- [13] Y. Qiu, T. L. Magnanti: Sensitivity analysis for variational inequalities, *Math. Oper. Res.* 17(1) (1992) 61–76.
- [14] R. T. Rockafellar: *Convex Analysis*, Princeton Mathematical Series 28, Princeton University Press, Princeton (1970).
- [15] R. T. Rockafellar: Maximal monotone relations and the second derivatives of non-smooth functions, *Ann. Inst. H. Poincaré Anal. Non Linéaire* 2(3) (1985) 167–184.
- [16] R. T. Rockafellar: Proto-differentiability of set-valued mappings and its applications in optimization, *Ann. Inst. H. Poincaré Anal. Non Linéaire* 6 (1989) 449–482.
- [17] A. Shapiro: Sensitivity analysis of parameterized variational inequalities, *Math. Oper. Res.* 30(1) (2005) 109–126.
- [18] Y.-L. Yu: On decomposing the proximal map, *Adv. Neural Information Processing Systems* 26 (2013) 91–99.