A Perturbation Framework for Convex Minimization and Monotone Inclusion Problems with Nonlinear Compositions*

Luis M. Briceño-Arias[1] and Patrick L. Combettes[2]

[1]Universidad Técnica Federico Santa María, Departamento de Matemática, Santiago, Chile
luis.briceno@usm.cl

[2]North Carolina State University, Department of Mathematics, Raleigh, USA
plc@math.ncsu.edu

Abstract. We introduce a framework based on Rockafellar's perturbation theory to analyze and solve general nonsmooth convex minimization and monotone inclusion problems involving nonlinearly composed functions as well as linear compositions. Such problems have been investigated only from a primal perspective and only for nonlinear compositions of smooth functions in finite-dimensional spaces in the absence of linear compositions. In the context of Banach spaces, the proposed perturbation analysis serves as a foundation for the construction of a dual problem and of a maximally monotone Kuhn–Tucker operator which is decomposable as the sum of simpler monotone operators. In the Hilbertian setting, this decomposition leads to a block-iterative primal-dual algorithm that fully splits all the components of the problem and appears to be the first proximal splitting algorithm for handling nonlinear composite problems. Various applications are discussed.

Keywords. Convex optimization, duality, monotone operator, nonlinear composition, perturbation theory, proximal method, splitting algorithm.

MSC2010 subject classification: 46N10, 47J20, 49M27, 49M29, 65K05, 90C25.

1 Introduction

This paper concerns the analysis and the numerical solution of optimization and monotone inclusion problems involving nonlinear compositions of convex functions. The novelty and difficulty of such formulations reside in the nonlinear composite terms and, in particular, in the design of a splitting mechanism that will fully decompose them towards a fine convex analysis of the problem and the development of efficient numerical methods. To isolate this difficulty, we first study the following simpler minimization formulation, where $\Gamma_0(\mathcal{X})$ designates the class of proper lower semicontinuous

*Contact author: P. L. Combettes. Email: plc@math.ncsu.edu. Phone: +1 919 515 2671. The work of L. M. Briceño-Arias was supported by Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for centers of excellence and FONDECYT 1190871 from ANID-Chile, and the work of P. L. Combettes was supported by the National Science Foundation under grant DMS-1818946.

convex functions from a Banach space $\mathcal{X}$ to $]-\infty,+\infty]$, sri the strong relative interior, and $\Box$ the infimal convolution operation (see Section 2 for notation).

Problem 1.1 Let $\mathcal{X}$ and $\mathcal{Y}$ be reflexive real Banach spaces, let $\phi \in \Gamma_0(\mathbb{R})$ be increasing and not constant, let $f \in \Gamma_0(\mathcal{X})$, let $g \in \Gamma_0(\mathcal{Y})$, let $\ell \in \Gamma_0(\mathcal{Y})$, let $L\colon \mathcal{X} \to \mathcal{Y}$ be linear and bounded, and let $h \in \Gamma_0(\mathcal{X})$. Set

$$\phi \circ f\colon \mathcal{X} \to \left]-\infty,+\infty\right] : x \mapsto \begin{cases} \phi\big(f(x)\big), & \text{if } f(x) \in \operatorname{dom}\phi; \\ +\infty, & \text{if } f(x) \notin \operatorname{dom}\phi \end{cases} \tag{1.1}$$

and suppose that the following hold:

[a] $(\exists\, z \in \operatorname{dom} f)\ f(z) \in \operatorname{int}\operatorname{dom}\phi$.

[b] $0 \in \operatorname{sri}\big(f^{-1}(\operatorname{dom}\phi) - \operatorname{dom} h\big)$.

[c] $0 \in \operatorname{sri}\big(\operatorname{dom} g^* - \operatorname{dom}\ell^*\big)$.

[d] $0 \in \operatorname{sri}\big(L(\operatorname{dom} h \cap f^{-1}(\operatorname{dom}\phi)) - \operatorname{dom} g - \operatorname{dom}\ell\big)$.

The goal is to

$$\underset{x\in\mathcal{X}}{\text{minimize}}\;\; (\phi\circ f)(x) + (g\,\Box\,\ell)(Lx) + h(x), \tag{1.2}$$

and the set of solutions is denoted by $\mathscr{P}$.

To motivate our investigation, let us consider a few notable special cases of Problem 1.1.

Example 1.2 In Problem 1.1 suppose that $\phi = \iota_{]-\infty,0]}$. Then (1.2) reduces to the constrained minimization problem

$$\underset{\substack{x\in\mathcal{X}\\ f(x)\leqslant 0}}{\text{minimize}}\;\; (g\,\Box\,\ell)(Lx) + h(x), \tag{1.3}$$

which is pervasive in nonlinear programming. For instance, suppose that $\mathcal{X} = \mathbb{R}^N$ and that $f = \|\cdot\|_p^p - \eta^p$, where $\eta \in \left]0,+\infty\right[$ and $p \in \left[1,+\infty\right[$. Then (1.3) becomes

$$\underset{\substack{x\in\mathbb{R}^N\\ \|x\|_p\leqslant \eta}}{\text{minimize}}\;\; (g\,\Box\,\ell)(Lx) + h(x). \tag{1.4}$$

An instance of (1.4) in the context of machine learning is found in [31, Section 4.1].

Example 1.3 Let $\theta \in \Gamma_0(\mathbb{R})$ be an increasing function such that $\operatorname{dom}\theta = \left]-\infty,\eta\right[$ for some $\eta \in \left[0,+\infty\right]$, $\lim_{\xi\uparrow\eta}\theta(\xi) = +\infty$, and $(\operatorname{rec}\theta)(1) > 0$. Let $\alpha\colon \left]0,+\infty\right[\to \left]0,+\infty\right[$ be such that $\lim_{\rho\downarrow 0}\alpha(\rho) = 0$ and $\underline{\lim}_{\rho\downarrow 0}\,\alpha(\rho)/\rho > 0$. In Problem 1.1, set $\mathcal{X} = \mathcal{Y} = \mathbb{R}^N$, $h = 0$, $\ell = \iota_{\{0\}}$, $L = \operatorname{Id}$, and given $\rho \in \left]0,+\infty\right[$, $\phi\colon \xi \mapsto \alpha(\rho)\theta(\xi/\rho)$. Then (1.2) becomes

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\;\; \alpha(\rho)\theta\big(f(x)/\rho\big) + g(x). \tag{1.5}$$

The asymptotic behavior of this family of penalty-barrier minimization problems as $\rho \downarrow 0$ is investigated in [6].

Example 1.4 In Problem 1.1 suppose that $\ell = \iota_{\{0\}}$ and $\phi = \theta \circ \max\{0, \cdot\}$, where $\theta \in \Gamma_0(\mathbb{R})$ is even with $\operatorname{Argmin} \theta = \{0\}$. Then $\phi = \theta \circ \max\{0, \cdot\}$ is an increasing function in $\Gamma_0(\mathbb{R})$ which is not constant and (1.2) reduces to

$$\underset{x \in \mathcal{X}}{\text{minimize}} \;\; \theta\big(\max\{0, f(x)\}\big) + g(Lx) + h(x). \tag{1.6}$$

For instance, if C is a nonempty closed convex subset of $\mathcal{X}$ and $f = d_C - \varepsilon$, where $\varepsilon \in [0, +\infty[$, we recover a scenario discussed in [25]. A special case is given by $f = d_C$ and $\theta \colon \xi \mapsto \ln(\rho) - \ln(\rho - \xi)$ if $\xi < \rho$; and $+\infty$ if $\xi \geqslant \rho$, where $\rho \in \,]0, +\infty[$. Thus, $\phi \circ f = \ln(\rho) - \ln(\rho - d_C)$ if $d_C < \rho$; and $+\infty$ if $d_C \geqslant \rho$, and it acts as a barrier keeping solutions at distance at most ρ from the set C.

Example 1.5 Let $p \in [1, +\infty[$ and let $(\mathcal{X}_i)_{i \in I}$ and $(\mathcal{Y}_k)_{k \in K}$ be finite families of reflexive real Banach spaces. For every $i \in I$, let C_i be a nonempty closed convex subset of $\mathcal{X}_i$ and, for every $k \in K$, let $g_k \in \Gamma_0(\mathcal{Y}_k)$ and let $L_{k,i} \colon \mathcal{X}_i \to \mathcal{Y}_k$ be a bounded linear operator. In Problem 1.1, set $\mathcal{X} = \bigoplus_{i \in I} \mathcal{X}_i$, $\mathcal{Y} = \bigoplus_{k \in K} \mathcal{Y}_k$, $f \colon (x_i)_{i \in I} \mapsto (\sum_{i \in I} d_{C_i}^p(x_i))^{1/p}$, $g \colon (y_k)_{k \in K} \mapsto \sum_{k \in K} g_k(y_k)$, $\ell = \iota_{\{0\}}$, $L \colon (x_i)_{i \in I} \mapsto (\sum_{i \in I} L_{k,i} x_i)_{k \in K}$, and $\phi = (\max\{0, \cdot\})^p$. Then (1.2) reduces to

$$\underset{(x_i)_{i \in I} \in \times_{i \in I} \mathcal{X}_i}{\text{minimize}} \;\; \sum_{i \in I} d_{C_i}^p(x_i) + \sum_{k \in K} g_k\Bigg(\sum_{i \in I} L_{k,i} x_i\Bigg) + h\big((x_i)_{i \in I}\big). \tag{1.7}$$

This formulation covers signal processing and location problems [11, 25, 39].

There is a vast literature on Problem 1.1 in the case of linear compositions, that is, when $\phi \colon \xi \mapsto \xi$. In this context, the duality theory goes back to [46] when $h = 0$ and $\ell = \iota_{\{0\}}$. On the algorithmic front, the primal-dual strategy adopted in [26] when h is smooth and ℓ is strongly convex, which originates in [12], is ultimately rooted in Fenchel–Rockafellar duality and boils down to finding a zero of an associated Kuhn–Tucker operator via monotone operator splitting methods. Further algorithmic developments along these lines in the linear composition setting can be found in [12, 21, 26, 29, 55]. Unfortunately this methodology is not extendible to Problem 1.1 in the presence of a nonlinear function ϕ. In the nonlinear setting, in terms of convex analysis, the conjugate and the subdifferential of $\phi \circ f$ in (1.1) are derived in [22, 23], building up on the work of [33, 34, 35, 53]; see also [19] for the Euclidean setting. Optimality conditions for the minimization of $\phi \circ f + h$ are established in [41, 42]. Duality theory for the minimization problem (1.1) does not seem to have been studied, even in the case when $h = 0$ and $\ell = \iota_{\{0\}}$. On the numerical side, in the finite-dimensional setting, with f smooth, $\mathcal{Y} = \mathcal{X}$, and $g = h = \ell^* = 0$, Problem 1.1 has been studied in [9] (in the case when f is vector-valued) by linearizing the objective function; see also [17, 18, 36, 40, 44] and the references therein for alternative approximations of f in this type of scenario. However, in the general setting of Problem 1.1, solution methods are not available.

To address the gaps identified above, we propose a methodology based on Rockafellar's perturbation theory. This general theory was initiated in [47] and further developed in [48, 50]; see also [32]. Specifically, we introduce for Problem 1.1 the perturbation function

$$\begin{aligned} F \colon \mathcal{X} \times \mathbb{R} \times \mathcal{Y} &\to \,]-\infty, +\infty] \\ (x, \xi, y) &\mapsto \begin{cases} \phi\big(f(x) + \xi\big) + (g \,\square\, \ell)(Lx + y) + h(x), & \text{if } \; f(x) + \xi \in \operatorname{dom} \phi; \\ +\infty, & \text{if } \; f(x) + \xi \notin \operatorname{dom} \phi. \end{cases} \end{aligned} \tag{1.8}$$

In this context, perturbation variables $\xi \in \mathbb{R}$ and $y \in \mathcal{Y}$ are associated to the nonlinear and the linear composition, respectively, and Problem 1.1 can be expressed as

$$\underset{x \in \mathcal{X}}{\text{minimize}} \;\; F(x, 0, 0). \tag{1.9}$$

By specializing the general theory of [50] to the perturbation function (1.8), we derive a dual for Problem 1.1 involving variables for each composition, together with a Kuhn–Tucker operator. We then show that this Kuhn–Tucker operator can be decomposed as a sum of elementary maximally monotone operators. In the Hilbertian setting, we derive their resolvents and propose proximal splitting algorithms that use ϕ, f, g, ℓ, h, and L separately to solve Problem 1.1 and its dual. In this endeavor, we leverage the fact that, while the proximity operator of $\phi \circ f$ is usually unknown, those of ϕ and f are often available. The resulting algorithms capture methods that were known in the case when $\phi\colon \xi \mapsto \xi$.

Next, we consider the broader setting of monotone inclusions. We first note that the minimization Problem 1.1 is a special case of the inclusion problem

$$\text{find } x \in \mathcal{X} \text{ such that } 0 \in \partial(\phi \circ f)(x) + \big(L^* \circ (B \,\square\, D) \circ L\big)x + Ax, \tag{1.10}$$

involving a potential term $\phi \circ f$ and the maximally monotone operators $A = \partial h$, $B = \partial g$, and $D = \partial \ell$, where $B \,\square\, D$ designates the parallel sum of B and D. For general maximally monotone operators this type of inclusion problem mixing potential and nonpotential terms appears in several areas, including game theory, saddle problems, evolution equations, variational inequalities, deep neural networks, and diffusion/convection problems in physics; see, e.g., [1, 2, 5, 12, 13, 27, 28, 30, 37, 57]. Building up on the tools developed for Problem 1.1, we shall address the following general form of (1.10).

Problem 1.6 Let $\mathcal{X}$ be a reflexive real Banach space, let I and K be disjoint finite sets, and let $(\beta, \chi) \in \left]0, +\infty\right[^2$. For every $i \in I$, let $\phi_i \in \Gamma_0(\mathbb{R})$ be increasing and not constant and let $f_i \in \Gamma_0(\mathcal{X})$ be such that $(\operatorname{int} \operatorname{dom} \phi_i) \cap f_i(\operatorname{dom} f_i) \neq \varnothing$. For every $k \in K$, let $\mathcal{Y}_k$ be a reflexive real Banach space, let $B_k\colon \mathcal{Y}_k \to 2^{\mathcal{Y}_k^*}$ and $D_k\colon \mathcal{Y}_k \to 2^{\mathcal{Y}_k^*}$ be maximally monotone, and let $L_k\colon \mathcal{X} \to \mathcal{Y}_k$ be linear and bounded. Furthermore, let $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be maximally monotone, let $C\colon \mathcal{X} \to \mathcal{X}^*$ be β-cocoercive, and let $Q\colon \mathcal{X} \to \mathcal{X}^*$ be monotone and χ-Lipschitzian. The goal is to

$$\text{find } x \in \mathcal{X} \text{ such that } 0 \in \sum_{i \in I} \partial(\phi_i \circ f_i)(x) + \sum_{k \in K} \big(L_k^* \circ (B_k \,\square\, D_k) \circ L_k\big)x + Ax + Cx + Qx. \tag{1.11}$$

An algorithmic framework will be proposed to solve Problem 1.6 in the Hilbertian setting and various applications will be discussed. In particular, we shall design a proximal algorithm to solve minimization problems involving sums of maxima of convex functions.

We present our notation and provide preliminary results in Section 2. In Section 3, the perturbation F introduced in (1.8) is employed to derive a dual for Problem 1.1 and its Kuhn–Tucker operator, which is shown to be maximally monotone. We then obtain primal-dual solutions as the zeros of this Kuhn–Tucker operator, which we decompose as the sum of elementary monotone operators. This decomposition is exploited in Section 4 to derive splitting algorithms that use ϕ, f, g, ℓ, h, and L separately to solve Problem 1.1 when all the Banach spaces are Hilbertian. Section 5 is devoted to the analysis and the numerical solution of Problem 1.6 based on the tools developed in Sections 3 and 4. A highlight of that section is a block-iterative algorithm that fully splits all the components of Problem 1.6. Even in its nonblock-iterative implementation, this appears to be the first proximal splitting algorithm for handling nonlinear composite problems. Applications of the proposed framework are presented in Section 6.

2 Notation and preliminary results

Let $(\mathcal{X}, \|\cdot\|)$ and $(\mathcal{Y}, \|\cdot\|)$ be reflexive real Banach spaces, let $\mathcal{X}^*$ and $\mathcal{Y}^*$ denote their respective topological duals, and let $\langle \cdot \mid \cdot \rangle$ denote the standard bilinear forms on $\mathcal{X} \times \mathcal{X}^*$ and $\mathcal{Y} \times \mathcal{Y}^*$. Throughout, we use the following notational conventions.

Notation 2.1 A generic vector in $\mathcal{X}$ is denoted by x and a generic vector in $\mathcal{X}^*$ is denoted by x^*. Bold symbols such as $\boldsymbol{z}$ denote elements in product spaces.

The symbol $\mathcal{X} \oplus \mathcal{Y}$ designates the standard product vector space $\mathcal{X} \times \mathcal{Y}$ equipped with the pairing

$$\big(\forall (x,y) \in \mathcal{X} \times \mathcal{Y}\big)\big(\forall (x^*,y^*) \in \mathcal{X}^* \times \mathcal{Y}^*\big) \quad \langle (x,y),(x^*,y^*)\rangle = \langle x, x^*\rangle + \langle y, y^*\rangle \tag{2.1}$$

and the norm

$$\big(\forall (x,y) \in \mathcal{X} \times \mathcal{Y}\big) \quad \|(x,y)\| = \sqrt{\|x\|^2 + \|y\|^2}. \tag{2.2}$$

The power set of $\mathcal{X}^*$ is denoted by $2^{\mathcal{X}^*}$. Let $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be a set-valued operator. We denote by $\operatorname{dom} A = \big\{x \in \mathcal{X} \mid Ax \neq \varnothing\big\}$ the domain of A, by $\operatorname{zer} A = \big\{x \in \mathcal{X} \mid 0 \in Ax\big\}$ the set of zeros of A, by $\operatorname{gra} A = \big\{(x,x^*) \in \mathcal{X} \times \mathcal{X}^* \mid x^* \in Ax\big\}$ the graph of A, and by A^{-1} the inverse of A, which has graph $\big\{(x^*,x) \in \mathcal{X}^* \times \mathcal{X} \mid x^* \in Ax\big\}$. The parallel sum of A and $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ is

$$A \,\square\, B = \big(A^{-1} + B^{-1}\big)^{-1}. \tag{2.3}$$

Moreover, A is monotone if

$$\big(\forall (x,x^*) \in \operatorname{gra} A\big)\big(\forall (y,y^*) \in \operatorname{gra} A\big) \quad \langle x - y, x^* - y^*\rangle \geqslant 0, \tag{2.4}$$

and maximally so if there exists no monotone operator $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ such that $\operatorname{gra} A \subset \operatorname{gra} B \neq \operatorname{gra} A$. If $\mathcal{X}$ is Hilbertian and A is maximally monotone, $J_A = (\mathrm{Id} + A)^{-1}$ is the resolvent of A, which is single-valued with $\operatorname{dom} J_A = \mathcal{X}$. If A is single-valued and $\beta \in \left]0,+\infty\right[$, A is β-cocoercive if

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad \langle x - y, Ax - Ay\rangle \geqslant \beta\|Ax - Ay\|^2. \tag{2.5}$$

Let $f\colon \mathcal{X} \to [-\infty,+\infty]$. Then f is proper if $-\infty \notin f(\mathcal{X}) \neq \{+\infty\}$, the domain of f is $\operatorname{dom} f = \big\{x \in \mathcal{X} \mid f(x) < +\infty\big\}$, and the set of minimizers of f is $\operatorname{Argmin} f = \big\{x \in \operatorname{dom} f \mid (\forall y \in \mathcal{X})\ f(x) \leqslant f(y)\big\}$. The conjugate of f is $f^*\colon x^* \mapsto \sup_{x\in\mathcal{X}}(\langle x, x^*\rangle - f(x))$. If f is proper, its subdifferential is

$$\partial f\colon \mathcal{X} \to 2^{\mathcal{X}^*} \colon x \mapsto \big\{x^* \in \mathcal{X}^* \mid (\forall y \in \operatorname{dom} f)\ \langle y - x, x^*\rangle + f(x) \leqslant f(y)\big\}, \tag{2.6}$$

its recession function is $\operatorname{rec} f\colon \mathcal{X} \to \left]-\infty,+\infty\right] \colon y \mapsto \sup_{x\in\operatorname{dom} f}(f(x+y) - f(x))$, and its perspective is

$$\widetilde{f}\colon \mathcal{X} \times \mathbb{R} \to \left]-\infty,+\infty\right] \colon (x,\xi) \mapsto \begin{cases} \xi\, f(x/\xi), & \text{if } \xi > 0;\\ (\operatorname{rec} f)(x), & \text{if } \xi = 0;\\ +\infty, & \text{otherwise.}\end{cases} \tag{2.7}$$

We set

$$(\forall \xi \in [0,+\infty[) \quad \xi \odot f = \begin{cases} \iota_{\overline{\operatorname{dom}} f}, & \text{if } \xi = 0;\\ \xi f, & \text{if } \xi > 0.\end{cases} \tag{2.8}$$

The infimal convolution of f with a proper function $\ell\colon \mathcal{X} \to \left]-\infty,+\infty\right]$ is

$$f \,\square\, \ell\colon \mathcal{X} \to [-\infty,+\infty] \colon x \mapsto \inf_{y\in\mathcal{X}} \big(f(y) + \ell(x - y)\big). \tag{2.9}$$

Let C be a convex subset of $\mathcal{X}$. The indicator function of C is denoted by ι_C, the support function of C by σ_C, the strong relative interior of C, i.e., the set of points $x \in C$ such that the cone generated by $-x + C$ is a closed vector subspace of $\mathcal{X}$, by $\operatorname{sri} C$, and the distance to C by d_C, i.e., $d_C \colon x \mapsto \inf_{y \in C} \|x - y\|$. The barrier cone of C is $\operatorname{bar} C = \{x^* \in \mathcal{X}^* \mid \sup \langle C, x^* \rangle < +\infty\}$ and the normal cone operator of C is $N_C = \partial \iota_C$. Now suppose that $\mathcal{X}$ is Hilbertian. If $f \in \Gamma_0(\mathcal{X})$, for every $x \in \mathcal{X}$, $\operatorname{prox}_f x$ denotes the unique minimizer of $f + \|\cdot - x\|^2/2$ and $\operatorname{prox}_f = (\mathrm{Id} + \partial f)^{-1} = J_{\partial f}$ is called the proximity operator of f and $\operatorname{prox}_f(\mathcal{X}) \subset \operatorname{dom} \partial f \subset \operatorname{dom} f$. If C is nonempty and closed, $\operatorname{proj}_C = \operatorname{prox}_{\iota_C}$ is the projection operator onto C. For background on convex analysis and monotone operators, see [8, 58].

The following facts will be required subsequently.

Lemma 2.2 *Let $\mathcal{X}$ be a reflexive real Banach space, let $f \in \Gamma_0(\mathcal{X})$, and let $\xi \in [0, +\infty[$. Then the following hold:*

(i) $\xi \odot f \in \Gamma_0(\mathcal{X})$.

(ii) $[\widetilde{f}(\cdot, \xi)]^* = \xi \odot f^*$ *and* $(\xi \odot f)^* = \widetilde{f^*}(\cdot, \xi)$.

(iii) *We have*

$$\partial(\xi \odot f) = \begin{cases} N_{\overline{\operatorname{dom} f}}, & \text{if } \xi = 0; \\ \xi \partial f, & \text{if } \xi \in \left]0, +\infty\right[. \end{cases} \tag{2.10}$$

(iv) *Suppose that $\mathcal{X}$ is Hilbertian. Then*

$$\operatorname{prox}_{\xi \odot f} = \begin{cases} \operatorname{proj}_{\overline{\operatorname{dom} f}}, & \text{if } \xi = 0; \\ \operatorname{prox}_{\xi f}, & \text{if } \xi \in \left]0, +\infty\right[. \end{cases} \tag{2.11}$$

Proof. (i)–(ii): See [45, Theorem 3E].

(iii)–(iv): Clear from (i) and (2.8). ☐

Lemma 2.3 [45, Theorem 3F] *Let $\mathcal{X}$ be a reflexive real Banach space, let $f \in \Gamma_0(\mathcal{X})$, and set $C = \{(x^*, \xi^*) \in \mathcal{X}^* \oplus \mathbb{R} \mid f^*(x^*) + \xi^* \leqslant 0\}$. Then the following hold:*

(i) $\widetilde{f} = \sigma_C$.

(ii) $\widetilde{f} \in \Gamma_0(\mathcal{X} \oplus \mathbb{R})$.

Lemma 2.4 *Let $\mathcal{X}$ be a reflexive real Banach space, let $f \colon \mathcal{X} \to \left]-\infty, +\infty\right]$ be proper, let $x \in \mathcal{X}$, and let $x^* \in \mathcal{X}^*$. Then the following hold:*

(i) $f(x) + f^*(x^*) \geqslant \langle x, x^* \rangle$. ([58, Theorem 2.3.1(ii)])

(ii) $f(x) + f^*(x^*) = \langle x, x^* \rangle \Leftrightarrow (x, x^*) \in \operatorname{gra} \partial f$. ([58, Theorem 2.4.2(iii)])

Lemma 2.5 *Let $\mathcal{X}$ be a reflexive real Banach space and let $f \in \Gamma_0(\mathcal{X})$. Then the following hold:*

(i) $f^* \in \Gamma_0(\mathcal{X}^*)$ *and* $f = f^{**}$. ([58, Theorem 2.3.3])

(ii) $(\partial f)^{-1} = \partial f^*$. ([58, Theorem 2.4.4(iv)])

(iii) ∂f *is maximally monotone.* ([58, Theorem 3.2.8])

(iv) $\operatorname{Argmin} f = \operatorname{zer} \partial f$. ([58, Theorem 2.5.7])

Lemma 2.6 *Let $\mathcal{X}$ be a reflexive real Banach space, let $f \in \Gamma_0(\mathcal{X})$, and let $g \in \Gamma_0(\mathcal{X})$. Suppose that* $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$. *Then the following hold:*

(i) $(f+g)^* = f^* \,\Box\, g^* \in \Gamma_0(\mathcal{X}^*)$. ([4, Theorem 1.1])

(ii) $\partial(f+g) = \partial f + \partial g$. ([4, Corollary 2.1])

Lemma 2.7 *Let $\mathcal{X}$ be a reflexive real Banach space, let $f \in \Gamma_0(\mathcal{X})$, and let $g \in \Gamma_0(\mathcal{X})$. Suppose that* $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$. *Then* $(\partial f) \,\Box\, (\partial g) = \partial(f \,\Box\, g)$.

Proof. It follows from (2.3), Lemma 2.5, and Lemma 2.6 applied to f^* and g^* that $(\partial f) \,\Box\, (\partial g) = ((\partial f)^{-1} + (\partial g)^{-1})^{-1} = (\partial f^* + \partial g^*)^{-1} = (\partial(f^* + g^*))^{-1} = \partial(f^* + g^*)^* = \partial(f^{**} \,\Box\, g^{**}) = \partial(f \,\Box\, g)$. ☐

Lemma 2.8 *Let $\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right]$ be an increasing proper convex function. Then the following hold:*

(i) $\operatorname{dom}\phi$ *is an interval and* $\inf \operatorname{dom}\phi = -\infty$.

(ii) $\operatorname{dom}\phi^* \subset \left[0, +\infty\right[$.

(iii) *Suppose that ϕ is lower semicontinuous and not constant. Then* $\operatorname{dom}\phi^* \cap \left]0, +\infty\right[\neq \varnothing$.

Proof. (i): Since ϕ is convex, $\operatorname{dom}\phi$ is convex, and hence an interval. Now take $\xi \in \operatorname{dom}\phi$ and $\eta \in \left]-\infty, \xi\right[$. Then $\phi(\eta) \leqslant \phi(\xi) < +\infty$ and therefore $\eta \in \operatorname{dom}\phi$. Consequently, $\inf \operatorname{dom}\phi = -\infty$.

(ii): Since ϕ is increasing, it follows from (i) that

$$(\forall \xi^* \in \left]-\infty, 0\right[) \quad \phi^*(\xi^*) = \sup_{\xi \in \operatorname{dom}\phi} \big(\xi\xi^* - \phi(\xi)\big) = +\infty, \tag{2.12}$$

which yields $\operatorname{dom}\phi^* \subset \left[0, +\infty\right[$.

(iii): If $\operatorname{dom}\phi^* = \{0\}$, then $\phi^* = \iota_{\{0\}} + \phi^*(0)$ and hence, since $\phi \in \Gamma_0(\mathbb{R})$, $\phi = \phi^{**} \equiv -\phi^*(0)$ is constant. Therefore, the claim follows from (ii). ☐

The next result discusses convex analytical properties of nonlinear compositions. It provides in particular a connection between the conjugate of $\phi \circ f$ and the marginal of a function that involves the perspective of f (see (2.7)).

Proposition 2.9 *Let $\mathcal{X}$ be a reflexive real Banach space, let $\phi \in \Gamma_0(\mathbb{R})$ be a nonconstant increasing function, and let $f \in \Gamma_0(\mathcal{X})$ be such that $(\operatorname{dom}\phi) \cap f(\operatorname{dom} f) \neq \varnothing$. Then the following hold:*

(i) $\operatorname{dom}(\phi \circ f) = f^{-1}(\operatorname{dom}\phi)$.

(ii) $\phi \circ f \in \Gamma_0(\mathcal{X})$.

(iii) *Suppose that there exists $z \in \mathcal{X}$ such that $f(z) \in \operatorname{int}\operatorname{dom}\phi$, and let $x^* \in \mathcal{X}^*$. Then*

$$(\phi \circ f)^*(x^*) = \min_{\xi^* \in \mathbb{R}} \big(\phi^*(\xi^*) + \widetilde{f^*}(x^*, \xi^*)\big). \tag{2.13}$$

(iv) *Let $x \in \operatorname{dom} f$. Then*

$$\bigcup_{\xi^* \in \partial\phi(f(x))} \partial(\xi^* \odot f)(x) \subset \partial(\phi \circ f)(x). \tag{2.14}$$

(v) *Suppose that there exists* $z \in \operatorname{dom} f$ *such that* $f(z) \in \operatorname{int}\operatorname{dom}\phi$, *and let* $x \in \operatorname{dom} f$. *Then*

$$\partial(\phi \circ f)(x) = \bigcup_{\xi^* \in \partial\phi(f(x))} \partial(\xi^* \odot f)(x). \tag{2.15}$$

(vi) *Suppose that there exists* $z \in \operatorname{dom} f$ *such that* $f(z) \in \operatorname{int}\operatorname{dom}\phi$, *let* $x \in \operatorname{dom} f$, *let* $x^* \in \mathcal{X}^*$, *and let* $\xi^* \in \mathbb{R}$. *Then the following are equivalent:*

(a) $(\phi \circ f)(x) + \phi^*(\xi^*) = (\xi^* \odot f)(x)$ *and* $(\xi^* \odot f)(x) + (\xi^* \odot f)^*(x^*) = \langle x, x^* \rangle$.

(b) $(\phi \circ f)(x) + \phi^*(\xi^*) + (\xi^* \odot f)^*(x^*) = \langle x, x^* \rangle$.

(c) $(\phi \circ f)(x) + (\phi \circ f)^*(x^*) = \langle x, x^* \rangle$ *and* $\phi^*(\xi^*) + (\xi^* \odot f)^*(x^*) = (\phi \circ f)^*(x^*)$.

(vii) *Suppose that there exists* $z \in \operatorname{dom} f$ *such that* $f(z) \in \operatorname{int}\operatorname{dom}\phi$, *let* $x \in \operatorname{dom} f$, *let* $x^* \in \mathcal{X}^*$, *and let* $\xi^* \in \mathbb{R}$. *Then*

$$\begin{cases} \xi^* \in \partial\phi\big(f(x)\big) \\ x^* \in \partial(\xi^* \odot f)(x) \end{cases} \quad \Leftrightarrow \quad \begin{cases} x^* \in \partial(\phi \circ f)(x) \\ \xi^* \in \operatorname{Argmin}\big(\phi^* + \widetilde{f^*}(x^*, \cdot)\big). \end{cases} \tag{2.16}$$

Proof. (i): See (1.1).

(ii): In view of (1.1), convexity is established as in [8, Proposition 8.21]. Since ϕ is not constant, we can take $(\xi, \mu) \in \operatorname{gra}\partial\phi$ such that $\mu \neq 0$. By (2.6),

$$(\forall \eta \in \mathbb{R}) \quad (\eta - \xi)\mu + \phi(\xi) \leqslant \phi(\eta). \tag{2.17}$$

In particular, for $\eta < \xi$, since ϕ is increasing, (2.17) yields $\mu > 0$. In turn, $\phi(\eta) \geqslant (\eta - \xi)\mu + \phi(\xi) \uparrow +\infty$ when $\eta \to +\infty$ and we deduce from [22, Proposition 3.7] that $\phi \circ f$ is lower semicontinuous. Finally, properness follows from (i).

(iii): It follows from [22, Proposition 4.11ii)], Lemma 2.8(ii), and Lemma 2.2(ii) that

$$\begin{aligned}(\phi \circ f)^*(x^*) &= \min_{\xi^* \in [0,+\infty[} \big(\phi^*(\xi^*) + (\xi^* \odot f)^*(x^*)\big) \\ &= \min_{\xi^* \in \mathbb{R}} \big(\phi^*(\xi^*) + \widetilde{f^*}(x^*, \xi^*)\big).\end{aligned} \tag{2.18}$$

(iv): This follows from Lemma 2.2(iii) and [22, Proposition 4.4].

(v): This follows from Lemma 2.2(iii) and [22, Proposition 4.11i)].

(vi)(b)⇔(vi)(a): Since $f(x) \in \mathbb{R}$, Lemma 2.4(i), Lemma 2.2(i), and Lemma 2.5(i) yield

$$\begin{aligned}&(\phi \circ f)(x) + \phi^*(\xi^*) + (\xi^* \odot f)^*(x^*) = \langle x, x^* \rangle \\ &\Leftrightarrow \quad (\xi^* \odot f)(x) = \xi^* f(x) \leqslant \phi^*(\xi^*) + \phi(f(x)) = \langle x, x^* \rangle - (\xi^* \odot f)^*(x^*) \leqslant (\xi^* \odot f)(x).\end{aligned} \tag{2.19}$$

(vi)(b)⇔(vi)(c): It follows from (iii) that

$$\begin{aligned}&(\phi \circ f)(x) + \phi^*(\xi^*) + (\xi^* \odot f)^*(x^*) = \langle x, x^* \rangle \\ &\Leftrightarrow \quad (\phi \circ f)^*(x^*) \leqslant \phi^*(\xi^*) + (\xi^* \odot f)^*(x^*) = \langle x, x^* \rangle - (\phi \circ f)(x) \leqslant (\phi \circ f)^*(x^*).\end{aligned} \tag{2.20}$$

(vii): Note that (iii) and Lemma 2.2(ii) imply that

$$(\phi \circ f)^*(x^*) = \min_{\eta^* \in \mathbb{R}} \big(\phi^*(\eta^*) + \widetilde{f^*}(x^*, \eta^*)\big) = \min_{\eta^* \in \mathbb{R}} \big(\phi^*(\eta^*) + (\eta^* \odot f)^*(x^*)\big). \tag{2.21}$$

Hence, Lemma 2.4(ii) and (vi) yield

$$\begin{cases}\xi^* \in \partial\phi\big(f(x)\big)\\ x^* \in \partial(\xi^*\odot f)(x)\end{cases} \quad\Leftrightarrow\quad \begin{cases}(\phi\circ f)(x) + \phi^*(\xi^*) = (\xi^*\odot f)(x)\\ (\xi^*\odot f)(x) + (\xi^*\odot f)^*(x^*) = \langle x, x^*\rangle\end{cases}$$
$$\Leftrightarrow\quad \begin{cases}(\phi\circ f)(x) + (\phi\circ f)^*(x^*) = \langle x, x^*\rangle\\ \phi^*(\xi^*) + (\xi^*\odot f)^*(x^*) = (\phi\circ f)^*(x^*)\end{cases}$$
$$\Leftrightarrow\quad \begin{cases}x^* \in \partial(\phi\circ f)(x)\\ \xi^* \in \text{Argmin}\,\big(\phi^* + \widetilde{f^*}(x^*,\cdot)\big),\end{cases} \tag{2.22}$$

which completes the proof. ☐

We now bring into play Rockafellar's perturbation theory.

Definition 2.10 ([50]) Let $\mathcal{X}$ and $\mathcal{U}$ be reflexive real Banach spaces, let $\varphi\colon \mathcal{X} \to \left]-\infty,+\infty\right]$ be a proper function, and consider the primal problem

$$\underset{x\in\mathcal{X}}{\text{minimize}}\;\; \varphi(x). \tag{2.23}$$

Let $\Phi\colon \mathcal{X}\oplus\mathcal{U} \to \left]-\infty,+\infty\right]$ be a perturbation of φ, i.e., $(\forall x\in\mathcal{X})\; \varphi(x) = \Phi(x,0)$. The Lagrangian is

$$\mathscr{L}_\Phi\colon \mathcal{X}\oplus\mathcal{U}^* \mapsto [-\infty,+\infty] : (x,u^*) \mapsto \inf_{u\in\mathcal{U}} \big(\Phi(x,u) - \langle u, u^*\rangle\big), \tag{2.24}$$

the dual problem is

$$\underset{u^*\in\mathcal{U}^*}{\text{minimize}}\;\; \sup_{x\in\mathcal{X}} \big(-\mathscr{L}_\Phi(x,u^*)\big), \tag{2.25}$$

and the Kuhn–Tucker operator is

$$\mathscr{K}_\Phi\colon \mathcal{X}\oplus\mathcal{U}^* \to 2^{\mathcal{X}^*\oplus\mathcal{U}}\colon (x,u^*) \mapsto \partial\big(\mathscr{L}_\Phi(\cdot,u^*)\big)(x) \times \partial\big(-\mathscr{L}_\Phi(x,\cdot)\big)(u^*). \tag{2.26}$$

Lemma 2.11 *Let $\mathcal{X}$ and $\mathcal{U}$ be reflexive real Banach spaces, let $\varphi \in \Gamma_0(\mathcal{X})$, let $\Phi \in \Gamma_0(\mathcal{X}\oplus\mathcal{U})$ be a perturbation of φ, and let $\mathscr{K}_\Phi$ be the Kuhn–Tucker operator of* (2.26)*. Then the following hold:*

(i) *Let $x\in\mathcal{X}$. Then $(-\mathscr{L}_\Phi(x,\cdot))^* = \Phi(x,\cdot)$.*

(ii) *$\mathscr{K}_\Phi$ is maximally monotone.*

(iii) *Suppose that $(x,u^*) \in \text{zer}\,\mathscr{K}_\Phi$. Then x solves* (2.23) *and u^* solves* (2.25).

Proof. (i): Let $u\in\mathcal{U}$. Then, appealing to (2.24) and Lemma 2.5(i), we obtain

$$\big(-\mathscr{L}_\Phi(x,\cdot)\big)^*(u) = \sup_{u^*\in\mathcal{U}^*} \Big(\langle u, u^*\rangle - \big(\Phi(x,\cdot)\big)^*(u^*)\Big) = \big(\Phi(x,\cdot)\big)^{**}(u) = \Phi(x,u). \tag{2.27}$$

(ii): Since $\Phi \in \Gamma_0(\mathcal{X}\oplus\mathcal{U})$, it follows from [50, Theorem 6 and Example 13] that the Lagrangian $\mathscr{L}_\Phi$ is a closed proper saddle function in the sense of [49, Section 3]. The claim therefore follows from [49, Theorem 3].

(iii): See [50, Theorem 15]. ☐

3 Perturbation theory for nonlinear composite minimization

Our strategy to analyze and solve Problem 1.1 hinges on the perturbation function F introduced in (1.8). We first study the nonlinear composition component of this perturbation.

Proposition 3.1 *Let $\mathcal{X}$ be a reflexive real Banach space, let $\phi \in \Gamma_0(\mathbb{R})$ be a nonconstant increasing function, let $f \in \Gamma_0(\mathcal{X})$ be such that $(\operatorname{dom}\phi) \cap f(\operatorname{dom} f) \neq \varnothing$, and let*

$$\begin{aligned}\Psi\colon \mathcal{X} \oplus \mathbb{R} &\to \left]-\infty,+\infty\right]\\ (x,\xi) &\mapsto \begin{cases}\phi\big(f(x)+\xi\big), & \text{if } f(x)+\xi \in \operatorname{dom}\phi;\\ +\infty, & \text{if } f(x)+\xi \notin \operatorname{dom}\phi\end{cases}\end{aligned} \tag{3.1}$$

be a perturbation of $\phi \circ f$. Let $\mathscr{L}_\Psi$ be the associated Lagrangian (see (2.24)) and let $\mathscr{K}_\Psi$ be the associated Kuhn–Tucker operator (see (2.26)). Then the following hold:

(i) $\Psi \in \Gamma_0(\mathcal{X} \oplus \mathbb{R})$.

(ii) *We have*

$$\begin{aligned}\mathscr{L}_\Psi\colon \mathcal{X} \oplus \mathbb{R} &\to [-\infty,+\infty]\\ (x,\xi^*) &\mapsto \begin{cases}+\infty, & \text{if } x \notin \operatorname{dom} f;\\ (\xi^* \odot f)(x) - \phi^*(\xi^*), & \text{if } x \in \operatorname{dom} f \text{ and } \xi^* \in \operatorname{dom}\phi^*;\\ -\infty, & \text{if } x \in \operatorname{dom} f \text{ and } \xi^* \notin \operatorname{dom}\phi^*.\end{cases}\end{aligned} \tag{3.2}$$

(iii) *We have*

$$\begin{aligned}\mathscr{K}_\Psi\colon \mathcal{X} \oplus \mathbb{R} &\to 2^{\mathcal{X}^* \oplus \mathbb{R}}\\ (x,\xi^*) &\mapsto \begin{cases}\partial(\xi^* \odot f)(x) \times \big(\partial\phi^*(\xi^*) - f(x)\big), & \text{if } x \in \operatorname{dom} f \text{ and } \xi^* \in \operatorname{dom}\phi^*;\\ \varnothing, & \text{if } x \notin \operatorname{dom} f \text{ or } \xi^* \notin \operatorname{dom}\phi^*.\end{cases}\end{aligned} \tag{3.3}$$

(iv) $\operatorname{dom}\mathscr{K}_\Psi \subset \operatorname{dom} f \times \left[0,+\infty\right[$.

(v) *$\mathscr{K}_\Psi$ is maximally monotone.*

Proof. (i): Set $\boldsymbol{f}\colon \mathcal{X} \oplus \mathbb{R} \to \left]-\infty,+\infty\right] : (x,\xi) \mapsto f(x)+\xi$. Then $\boldsymbol{f} \in \Gamma_0(\mathcal{X} \oplus \mathbb{R})$ and (1.1) implies that $\Psi = \phi \circ \boldsymbol{f}$. Thus, since $\varnothing \neq (\operatorname{dom}\phi) \cap f(\operatorname{dom} f) \subset (\operatorname{dom}\phi) \cap \boldsymbol{f}(\operatorname{dom}\boldsymbol{f})$, the result follows from Proposition 2.9(ii).

(ii): It follows from (2.24) that, for every $(x,\xi^*) \in \mathcal{X} \times \mathbb{R}$,

$$\begin{aligned}\mathscr{L}_\Psi(x,\xi^*) &= \inf_{\xi\in\mathbb{R}} \big(\Psi(x,\xi) - \xi\xi^*\big)\\ &= \begin{cases}-\displaystyle\sup_{\xi\in(\operatorname{dom}\phi)-f(x)} \Big(\xi\xi^* - \phi\big(f(x)+\xi\big)\Big), & \text{if } x \in \operatorname{dom} f;\\ +\infty, & \text{if } x \notin \operatorname{dom} f;\end{cases}\\ &= \begin{cases}\xi^* f(x) - \displaystyle\sup_{\xi+f(x)\in\operatorname{dom}\phi} \Big(\big(\xi+f(x)\big)\xi^* - \phi\big(f(x)+\xi\big)\Big), & \text{if } x \in \operatorname{dom} f;\\ +\infty, & \text{if } x \notin \operatorname{dom} f;\end{cases}\end{aligned}$$

$$= \begin{cases} \xi^* f(x) - \phi^*(\xi^*), & \text{if } x \in \operatorname{dom} f; \\ +\infty, & \text{if } x \notin \operatorname{dom} f. \end{cases} \tag{3.4}$$

In view of Lemma 2.8(ii) and (2.8), we obtain (3.2).

(iii): Let $x \in \mathcal{X}$ and $\xi^* \in \mathbb{R}$. If $x \notin \operatorname{dom} f$, then (ii) yields $\mathscr{L}_\Psi(x, \xi^*) = +\infty$ and hence $\mathscr{K}_\Psi(x, \xi^*) = \varnothing$ in view of (2.26). Similarly, if $\xi^* \notin \operatorname{dom} \phi^*$, then $-\mathscr{L}_\Psi(x, \xi^*) = +\infty$ and hence $\mathscr{K}_\Psi(x, \xi^*) = \varnothing$. Now suppose that $x \in \operatorname{dom} f$ and $\xi^* \in \operatorname{dom} \phi^*$. Then it follows from (ii) and Lemma 2.8(ii) that $\partial(\mathscr{L}_\Psi(\cdot, \xi^*)) = \partial(\xi^* \odot f)$. Moreover, for every $\xi \in \mathbb{R}$, we derive from Lemma 2.4(ii), Lemma 2.11(i), (ii), Lemma 2.5(i), and (3.1) that

$$\begin{aligned} \xi \in \partial\big(-\mathscr{L}_\Psi(x, \cdot)\big)(\xi^*) \quad &\Leftrightarrow \quad -\mathscr{L}_\Psi(x, \xi^*) + \big(-\mathscr{L}_\Psi(x, \cdot)\big)^*(\xi) = \xi\xi^* \\ &\Leftrightarrow \quad \phi^*(\xi^*) + \phi\big(f(x) + \xi\big) = \big(f(x) + \xi\big)\xi^* \\ &\Leftrightarrow \quad f(x) + \xi \in \partial\phi^*(\xi^*). \end{aligned} \tag{3.5}$$

Altogether, this verifies that (2.26) assumes the form announced in (3.3).

(iv): It follows from (3.3) and Lemma 2.8(ii) that $\operatorname{dom} \mathscr{K}_\Psi \subset \operatorname{dom} f \times \operatorname{dom} \partial\phi^* \subset \operatorname{dom} f \times \operatorname{dom} \phi^* \subset \operatorname{dom} f \times [0, +\infty[$.

(v): This is a consequence of (i) and Lemma 2.11(ii). ☐

Next, we provide a characterization of the solutions to Problem 1.1 in terms of a monotone inclusion problem in $\mathcal{X} \oplus \mathbb{R}$.

Proposition 3.2 *Consider the setting of Problem 1.1 and the Kuhn–Tucker operator $\mathscr{K}_\Psi$ of Proposition 3.1(iii). Set*

$$\boldsymbol{B} \colon \mathcal{X} \oplus \mathbb{R} \to 2^{\mathcal{X}^* \oplus \mathbb{R}} \colon (x, \xi^*) \mapsto \Big(\big(L^* \circ \big((\partial g) \,\Box\, (\partial \ell)\big) \circ L\big)(x) + \partial h(x)\Big) \times \{0\}. \tag{3.6}$$

Then the following hold:

(i) *$\boldsymbol{B}$ is maximally monotone.*

(ii) $\mathscr{P} = \bigcup_{\xi^* \in \mathbb{R}} \big\{x \in \mathcal{X} \mid (x, \xi^*) \in \operatorname{zer}(\mathscr{K}_\Psi + \boldsymbol{B})\big\}$.

Proof. Lemma 2.5(i), Lemma 2.6(i), and assumption [c] in Problem 1.1 yield

$$g \,\Box\, \ell \in \Gamma_0(\mathcal{Y}). \tag{3.7}$$

(i): Since assumption [d] in Problem 1.1 implies that $0 \in \operatorname{sri}(L(\operatorname{dom} h) - \operatorname{dom}(g \,\Box\, \ell))$, we derive from (3.7) that $(g \,\Box\, \ell) \circ L + h \in \Gamma_0(\mathcal{X})$ and it follows from Lemma 2.7, [58, Theorem 2.8.3(vii)], and Lemma 2.5(iii) that

$$L^* \circ \big((\partial g) \,\Box\, (\partial \ell)\big) \circ L + \partial h = L^* \circ \partial(g \,\Box\, \ell) \circ L + \partial h = \partial\big((g \,\Box\, \ell) \circ L + h\big) \tag{3.8}$$

is maximally monotone. Thus, $\boldsymbol{B}$ is maximally monotone.

(ii): It follows from assumption [a] in Problem 1.1 and Proposition 2.9(ii) that $\phi \circ f \in \Gamma_0(\mathcal{X})$, and therefore from assumption [b] that $(\phi \circ f) + h \in \Gamma_0(\mathcal{X})$. Altogether, using Lemma 2.5(iv), (3.7), assumptions [a]–[d], [58, Theorem 2.8.3(vii)], Proposition 2.9(v), Lemma 2.7, and Lemma 2.5(ii), we obtain

$$x \in \mathscr{P} \qquad \Leftrightarrow \qquad 0 \in \partial\big(\phi \circ f + (g \,\Box\, \ell) \circ L + h\big)(x)$$

$$
\begin{aligned}
&\Leftrightarrow \quad \big(\exists\, \xi^* \in \partial\phi\big(f(x)\big)\big) \quad 0 \in \partial(\xi^* \odot f)(x) + L^*\big(\partial(g \,\square\, \ell)(Lx)\big) + \partial h(x) \\
&\Leftrightarrow \quad (\exists\, \xi^* \in \mathbb{R}) \quad \begin{cases} 0 \in \partial(\xi^* \odot f)(x) + L^*\Big(\big((\partial g) \,\square\, (\partial\ell)\big)(Lx)\Big) + \partial h(x) \\ 0 \in \partial\phi^*(\xi^*) - f(x). \end{cases}
\end{aligned} \tag{3.9}
$$

Thus, the result follows from Proposition 3.1(iii) and (3.6). ☐

Following the abstract perturbation theory of [50] outlined in Definition 2.10, the perturbation (1.8) allows us to construct a Lagrangian for Problem 1.1, a dual problem, and a Kuhn–Tucker operator that will provide solutions to Problem 1.1 and its dual. This program is described in the next theorem.

Theorem 3.3 *Consider the setting of Problem 1.1, set $D = \operatorname{dom} f \cap \operatorname{dom} h$ and set $V = \operatorname{dom} \phi^* \times (\operatorname{dom} g^* \cap \operatorname{dom} \ell^*)$. Then the following hold with respect to the perturbation function F of (1.8):*

(i) *The Lagrangian is*

$$
\mathscr{L}_F \colon \mathcal{X} \times \mathbb{R} \times \mathcal{Y}^* \to [-\infty, +\infty]
$$
$$
(x, \xi^*, y^*) \mapsto \begin{cases} +\infty, & \text{if } x \notin D; \\ (\xi^* \odot f)(x) + h(x) + \langle Lx, y^*\rangle - \phi^*(\xi^*) - g^*(y^*) - \ell^*(y^*), & \\ & \text{if } x \in D \text{ and } (\xi^*, y^*) \in V; \\ -\infty, & \text{if } x \in D \text{ and } (\xi^*, y^*) \notin V. \end{cases} \tag{3.10}
$$

(ii) *The dual problem is*

$$
\underset{(\xi^*, y^*) \in \mathbb{R} \times \mathcal{Y}^*}{\text{minimize}} \quad \phi^*(\xi^*) + \big(h^* \,\square\, \widetilde{f^*}(\cdot, \xi^*)\big)(-L^* y^*) + g^*(y^*) + \ell^*(y^*). \tag{3.11}
$$

(iii) *The Kuhn–Tucker operator is*

$$
\mathscr{K}_F \colon \mathcal{X} \times \mathbb{R} \times \mathcal{Y}^* \to 2^{\mathcal{X}^* \times \mathbb{R} \times \mathcal{Y}}
$$
$$
(x, \xi^*, y^*) \mapsto \begin{cases} \big(\partial(\xi^* \odot f)(x) + \partial h(x) + L^* y^*\big) \times \big(\partial\phi^*(\xi^*) - f(x)\big) & \\ \quad \times \big(\partial g^*(y^*) + \partial\ell^*(y^*) - Lx\big), & \text{if } x \in D \text{ and } (\xi^*, y^*) \in V; \\ \varnothing, & \text{if } x \notin D \text{ or } (\xi^*, y^*) \notin V. \end{cases} \tag{3.12}
$$

(iv) *Let $\mathscr{D}$ be the set of solutions to (3.11). Then $\operatorname{zer} \mathscr{K}_F \subset \mathscr{P} \times \mathscr{D}$.*

(v) $\mathscr{P} = \displaystyle\bigcup_{(\xi^*, y^*) \in \mathbb{R} \times \mathcal{Y}^*} \big\{ x \in \mathcal{X} \mid (x, \xi^*, y^*) \in \operatorname{zer} \mathscr{K}_F \big\}$.

Proof. Note that assumptions [a]–[d] imply that $D = \operatorname{dom} f \cap \operatorname{dom} h \neq \varnothing$ and $V = \operatorname{dom} \phi^* \times (\operatorname{dom} g^* \cap \operatorname{dom} \ell^*) \neq \varnothing$.

(i): Let $(x, \xi^*, y^*) \in \mathcal{X} \times \mathbb{R} \times \mathcal{Y}^*$. In view of (2.24),

$$
\mathscr{L}_F(x, \xi^*, y^*) = \inf_{(\xi, y) \in \mathbb{R} \oplus \mathcal{Y}} \big(F(x, \xi, y) - \xi\xi^* - \langle y, y^*\rangle \big). \tag{3.13}
$$

If $x \notin D$, (1.8) implies that $F(x, \xi, y) = +\infty$ and, therefore, that $\mathscr{L}_F(x, \xi^*, y^*) = +\infty$. Now suppose that $x \in D$. Then (3.13) and (1.8) yield

$$
\mathscr{L}_F(x, \xi^*, y^*) = h(x) + \inf_{\xi \in (\operatorname{dom} \phi) - f(x)} \big(\phi(f(x) + \xi) - \xi\xi^*\big) + \inf_{y \in \mathcal{Y}} \big((g \,\square\, \ell)(Lx + y) - \langle y, y^*\rangle\big)
$$

$$
\begin{aligned}
&= \xi^* f(x) + h(x) + \langle Lx, y^*\rangle - \sup_{f(x)+\xi\in\operatorname{dom}\phi} \big((f(x)+\xi)\xi^* - \phi(f(x)+\xi)\big)\\
&\qquad - \sup_{y\in\mathcal{Y}} \big(\langle Lx+y, y^*\rangle - (g\,\square\,\ell)(Lx+y)\big)\\
&= (\xi^* \odot f)(x) + h(x) + \langle Lx, y^*\rangle - \phi^*(\xi^*) - (g\,\square\,\ell)^*(y^*)\\
&= (\xi^* \odot f)(x) + h(x) + \langle Lx, y^*\rangle - \phi^*(\xi^*) - g^*(y^*) - \ell^*(y^*). \qquad (3.14)
\end{aligned}
$$

(ii): First, for every $(\xi^*, y^*) \in (\mathbb{R}\times\mathcal{Y}^*) \smallsetminus V$, we derive from (i) that $\sup_{x\in\mathcal{X}} \big(-\mathscr{L}_F(x,\xi^*,y^*)\big) = +\infty$. Next, for every $(\xi^*, y^*) \in V$, we derive from Lemma 2.6(i) and Lemma 2.2(ii) that

$$
\begin{aligned}
\sup_{x\in\mathcal{X}} \big(-\mathscr{L}_F(x,\xi^*,y^*)\big) &= \phi^*(\xi^*) + g^*(y^*) + \ell^*(y^*) + \sup_{x\in D} \big(\langle x, -L^*y^*\rangle - (h+\xi^* f)(x)\big)\\
&= \phi^*(\xi^*) + g^*(y^*) + \ell^*(y^*) + (h+\xi^*\odot f)^*(-L^*y^*)\\
&= \phi^*(\xi^*) + g^*(y^*) + \ell^*(y^*) + \big(h^*\,\square\,\widetilde{f^*}(\cdot,\xi^*)\big)(-L^*y^*). \qquad (3.15)
\end{aligned}
$$

According to (2.25), the dual problem is

$$
\underset{(\xi^*,y^*)\in\mathbb{R}\times\mathcal{Y}^*}{\text{minimize}} \;\; \sup_{x\in\mathcal{X}} \big(-\mathscr{L}_F(x,\xi^*,y^*)\big) \qquad (3.16)
$$

which, in view of (3.15), is precisely (3.11).

(iii): Let $(x,\xi^*,y^*) \in \mathcal{X}\times\mathbb{R}\times\mathcal{Y}^*$. If $x \notin D$, then it results from (i) that $\mathscr{L}_F(x,\xi^*,y^*) = +\infty$ and therefore $\mathscr{K}_F(x,\xi^*,y^*) = \varnothing$. If $(\xi^*,y^*) \notin V$, then $-\mathscr{L}_F(x,\xi^*,y^*) = +\infty$ and therefore $\mathscr{K}_F(x,\xi^*,y^*) = \varnothing$. Now suppose that $x \in D$ and $(\xi^*,y^*) \in V$, and note that

$$
\mathscr{L}_F(x,\xi^*,y^*) = \mathscr{L}_\Psi(x,\xi^*) + h(x) + \langle x, L^*y^*\rangle - g^*(y^*) - \ell^*(y^*), \qquad (3.17)
$$

where $\mathscr{L}_\Psi$ is defined in Proposition 3.1(ii). Therefore, assumption [b] and Lemma 2.6(ii) yield

$$
\partial\big(\mathscr{L}_F(\cdot,\xi^*,y^*)\big) = \partial\big(\mathscr{L}_\Psi(\cdot,\xi^*)\big) + \partial h + L^*y^*, \qquad (3.18)
$$

while assumption [c] and Lemma 2.6(ii) imply that

$$
\partial\big(-\mathscr{L}_F(x,\cdot,\cdot)\big) \colon (\xi^*,y^*) \mapsto \partial\big(-\mathscr{L}_\Psi(x,\cdot)\big)(\xi^*) \times \big(\partial g^*(y^*) + \partial\ell^*(y^*) - Lx\big). \qquad (3.19)
$$

Thus, (3.12) follows from (2.26) and Proposition 3.1(iii).

(iv): This follows from (iii) and Lemma 2.11(iii).

(v): Suppose that $x \in \mathscr{P}$. Then it follows from Proposition 3.2(ii) that there exists $\xi^* \in \mathbb{R}$ such that

$$
\begin{cases}
0 \in \partial(\xi^*\odot f)(x) + L^*\Big(\big((\partial g)\,\square\,(\partial\ell)\big)(Lx)\Big) + \partial h(x)\\
0 \in \partial\phi^*(\xi^*) - f(x).
\end{cases} \qquad (3.20)
$$

Thus, Lemma 2.5(ii) and (2.3) guarantee the existence of $(\xi^*,y^*) \in \mathbb{R}\times\mathcal{Y}^*$ such that

$$
\begin{cases}
0 \in \partial(\xi^*\odot f)(x) + \partial h(x) + L^*y^*\\
0 \in \partial\phi^*(\xi^*) - f(x)\\
0 \in \partial g^*(y^*) + \partial\ell^*(y^*) - Lx,
\end{cases} \qquad (3.21)
$$

which shows that $(x,\xi^*,y^*) \in \operatorname{zer}\mathscr{K}_F$. Conversely, if $(x,\xi^*,y^*) \in \operatorname{zer}\mathscr{K}_F$, then (iii) yields $x \in \mathscr{P}$. ☐

Remark 3.4 In Theorem 3.3, suppose that $\phi\colon \xi \mapsto \xi$, $h = 0$, and $\ell = \iota_{\{0\}}$. Then $\phi^* = \iota_{\{1\}}$, the primal problem (1.2) reduces to

$$\underset{x\in\mathcal{X}}{\text{minimize}}\;\; f(x) + g(Lx), \tag{3.22}$$

the perturbation function (1.8) to $(x, y) \mapsto f(x) + g(Lx + y)$, the Lagrangian (3.10) to

$$(x, y^*) \mapsto \begin{cases} +\infty, & \text{if } x \notin \operatorname{dom} f;\\ f(x) + \langle Lx, y^*\rangle - g^*(y^*), & \text{if } x \in \operatorname{dom} f \text{ and } y^* \in \operatorname{dom} g^*;\\ -\infty, & \text{if } x \in \operatorname{dom} f \text{ and } y^* \notin \operatorname{dom} g^*, \end{cases} \tag{3.23}$$

the dual problem (3.11) to

$$\underset{y^*\in\mathcal{Y}^*}{\text{minimize}}\;\; f^*(-L^*y^*) + g^*(y^*), \tag{3.24}$$

and the Kuhn–Tucker operator (3.12) to $(x, y^*) \mapsto \big(\partial f(x) + L^*y^*\big) \times \big(\partial g^*(y^*) - Lx\big)$. We thus recover the classical Lagrangian, Fenchel–Rockafellar dual, and Kuhn–Tucker operator associated to (3.22) [50, Examples 11 & 11'].

Remark 3.5 In Theorem 3.3, suppose that $g = \ell = 0$. Then (1.2) reduces to

$$\underset{x\in\mathcal{X}}{\text{minimize}}\;\; (\phi \circ f)(x) + h(x) \tag{3.25}$$

and the perturbation function (1.8) becomes

$$(x, \xi) \mapsto \begin{cases} \phi(f(x) + \xi) + h(x), & \text{if } f(x) + \xi \in \operatorname{dom}\phi;\\ +\infty, & \text{if } f(x) + \xi \notin \operatorname{dom}\phi. \end{cases} \tag{3.26}$$

In turn, the Lagrangian (3.10) is

$$(x, \xi^*) \mapsto \begin{cases} +\infty, & \text{if } x \notin \operatorname{dom} f \;\text{ or }\; x \notin \operatorname{dom} h;\\ h(x) + (\xi^* \odot f)(x) - \phi^*(\xi^*), & \text{if } x \in \operatorname{dom} f \cap \operatorname{dom} h \text{ and } \xi^* \in \operatorname{dom}\phi^*;\\ -\infty, & \text{if } x \in \operatorname{dom} f \cap \operatorname{dom} h \text{ and } \xi^* \notin \operatorname{dom}\phi^*, \end{cases} \tag{3.27}$$

the dual problem (3.11) is

$$\underset{\xi^*\in\mathbb{R}}{\text{minimize}}\;\; \phi^*(\xi^*) + \big(h^* \,\square\, \widetilde{f^*}(\cdot, \xi^*)\big)(0), \tag{3.28}$$

and the Kuhn–Tucker operator (3.12) is

$$(x, \xi^*) \mapsto \begin{cases} \big(\partial(\xi^* \odot f)(x) + \partial h(x)\big) \times \big(\partial\phi^*(\xi^*) - f(x)\big), & \text{if } x \in \operatorname{dom} f \cap \operatorname{dom} h \;\text{ and }\; \xi^* \in \operatorname{dom}\phi^*;\\ \varnothing, & \text{if } x \notin \operatorname{dom} f \cap \operatorname{dom} h \;\text{ or }\; \xi^* \notin \operatorname{dom}\phi^*. \end{cases} \tag{3.29}$$

The perturbation function (3.26) appears in [52, Example 11.46] and [22]. It is employed in [10, Section I.4] to obtain (3.28) and in [51] in the context of augmented Lagrangian formulations.

Our next topic is the splitting of the Kuhn–Tucker operator of Theorem 3.3(iii) into elementary components. This decomposition will pave the way to the numerical solution of Problem 1.1.

Proposition 3.6 *Consider the setting of Problem 1.1, set* $\boldsymbol{\mathcal{X}} = \mathcal{X} \oplus \mathbb{R} \oplus \mathcal{Y}^*$*, let* $\mathscr{K}_\Psi$ *be the Kuhn–Tucker operator of Proposition 3.1(iii), and let* $\mathscr{K}_F$ *be the Kuhn–Tucker operator of Theorem 3.3(iii). Define*

$$\begin{cases} \boldsymbol{M}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}^*}\colon (x,\xi^*,y^*) \mapsto \mathscr{K}_\Psi(x,\xi^*) \times \partial g^*(y^*) \\ \boldsymbol{S}\colon \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}^*\colon (x,\xi^*,y^*) \mapsto (L^*y^*, 0, -Lx) \\ \boldsymbol{G}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}^*}\colon (x,\xi^*,y^*) \mapsto \partial h(x) \times \{0\} \times \partial \ell^*(y^*). \end{cases} \tag{3.30}$$

Then the following hold:

(i) $\mathscr{K}_F = \boldsymbol{M} + \boldsymbol{S} + \boldsymbol{G}$.

(ii) $\boldsymbol{M}$ *is maximally monotone.*

(iii) $\boldsymbol{S}$ *is bounded, linear, and skew in the sense that* $(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{X}})\ \langle \boldsymbol{x}, \boldsymbol{S}\boldsymbol{x}\rangle = 0$*. In addition,* $\|\boldsymbol{S}\| = \|L\|$.

(iv) $\boldsymbol{G}$ *is maximally monotone.*

(v) *Suppose that* h *and* ℓ^* *are differentiable and that their gradients are Lipschitzian with constants* $\delta \in \left]0,+\infty\right[$ *and* $\vartheta \in \left]0,+\infty\right[$*, respectively. Then* $\boldsymbol{G}$ *is* $\max\{\delta,\vartheta\}$*-Lipschitzian and* $\min\{1/\delta,1/\vartheta\}$*-cocoercive.*

Proof. Define

$$\boldsymbol{h}\colon \boldsymbol{\mathcal{X}} \to \left]-\infty,+\infty\right]\colon (x,\xi^*,y^*) \mapsto h(x) + \ell^*(y^*). \tag{3.31}$$

(i): Combine (3.3), (3.12), and (3.30).

(ii): As seen in Proposition 3.1(v), $\mathscr{K}_\Psi$ is maximally monotone. Therefore, it follows from Lemma 2.5(i) and Lemma 2.5(iii) that $\boldsymbol{M}$ is maximally monotone.

(iii): The first three claims are clear. Now let $\boldsymbol{x} = (x,\xi^*,y^*) \in \boldsymbol{\mathcal{X}}$. Then (2.2) yields

$$\|\boldsymbol{S}\boldsymbol{x}\|^2 = \|L^*y^*\|^2 + \|Lx\|^2 \leqslant \|L\|^2\|\boldsymbol{x}\|^2, \tag{3.32}$$

which shows that $\|\boldsymbol{S}\| \leqslant \|L\|$. On the other hand, suppose that $\|x\| \leqslant 1$ and that $(\xi^*,y^*) = (0,0)$. Then $\|Lx\| = \|\boldsymbol{S}\boldsymbol{x}\| \leqslant \|\boldsymbol{S}\|$ and, therefore, $\|L\| \leqslant \|\boldsymbol{S}\|$.

(iv): Since $\boldsymbol{h} \in \Gamma_0(\boldsymbol{\mathcal{X}})$, Lemma 2.5(iii) asserts that $\boldsymbol{G} = \partial\boldsymbol{h}$ is maximally monotone.

(v): By (3.31), $\boldsymbol{h}$ is differentiable and $\boldsymbol{G} = \nabla\boldsymbol{h}\colon (x,\xi^*,y^*) \mapsto (\nabla h(x), 0, \nabla\ell^*(y^*))$ is $\max\{\delta,\vartheta\}$-Lipschitzian, which makes it $\min\{1/\delta,1/\vartheta\}$-cocoercive by [7, Corollaire 10]. ☐

4 Proximal analysis and solution methods

We turn our attention to the design of algorithms for solving the primal problem (1.2) and its dual (3.11) in the case when $\mathcal{X}$ and $\mathcal{Y}$ are real Hilbert spaces. Theorem 3.3(iv) opens a path towards this objective by seeking a zero of the Kuhn–Tucker operator $\mathscr{K}_F$ of (3.12), i.e., in view of Proposition 3.6(i), of the sum of the maximally monotone operators $\boldsymbol{M}$, $\boldsymbol{S}$, and $\boldsymbol{G}$. This can be achieved by splitting methods which activate these operators separately and involve resolvents.

4.1 Resolvent computations

We start with the computation of the resolvent of the Kuhn–Tucker operator $\mathscr{K}_\Psi$.

Proposition 4.1 *Let $\mathcal{X}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{X})$, and let $\phi \in \Gamma_0(\mathbb{R})$ be a nonconstant increasing function such that $(\operatorname{int}\operatorname{dom}\phi) \cap f(\operatorname{dom} f) \neq \varnothing$. Let $\mathscr{K}_\Psi$ be the Kuhn–Tucker operator defined in Proposition 3.1(iii), let $x \in \mathcal{X}$, let $\xi^* \in \mathbb{R}$, and let $\gamma \in \left]0,+\infty\right[$. Then there exists a unique real number $\omega \in \left[0,+\infty\right[$ such that*

$$\omega = \operatorname{prox}_{\gamma\phi^*}\big(\xi^* + \gamma f(\operatorname{prox}_{\gamma(\omega\odot f)} x)\big). \tag{4.1}$$

Moreover,

$$J_{\gamma\mathscr{K}_\Psi}(x,\xi^*) = \big(\operatorname{prox}_{\gamma(\omega\odot f)} x, \omega\big). \tag{4.2}$$

Proof. We deduce from Proposition 3.1(v) and [8, Proposition 23.8] that $J_{\mathscr{K}_\Psi}$ is single-valued on $\mathcal{X}\times\mathbb{R}$. Now let $(p,\omega) \in \mathcal{X}\times\mathbb{R}$. Then, by (3.3), Proposition 3.1(iv), and Lemma 2.2(i),

$$\begin{aligned}
(p,\omega) = J_{\gamma\mathscr{K}_\Psi}(x,\xi^*) \quad&\Leftrightarrow\quad (x,\xi^*) \in (p,\omega) + \gamma\mathscr{K}_\Psi(p,\omega)\\
&\Leftrightarrow\quad \begin{cases} p \in \operatorname{dom} f \text{ and } \omega \in \left[0,+\infty\right[\\ x \in p + \gamma\partial(\omega\odot f)(p)\\ \xi^* \in \omega + \gamma\partial\phi^*(\omega) - \gamma f(p)\end{cases}\\
&\Leftrightarrow\quad \begin{cases} p \in \operatorname{dom} f \text{ and } \omega \in \left[0,+\infty\right[\\ p = \operatorname{prox}_{\gamma(\omega\odot f)} x\\ \omega = \operatorname{prox}_{\gamma\phi^*}\big(\xi^* + \gamma f(p)\big),\end{cases}
\end{aligned} \tag{4.3}$$

which proves the assertion. ☐

Remark 4.2 In the setting of Proposition 4.1, let us provide details on the computation of the unique solution $\omega \in \left[0,+\infty\right[$ to (4.1) and of $J_{\gamma\mathscr{K}_\Psi}(x,\xi^*)$. These computations can be performed by testing the membership of $\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom}} f} x)$ in $\operatorname{Argmin}\phi$. Using [8, Proposition 12.29 and Theorem 14.3(ii)] as well as Lemma 2.2(iv), we first observe that

$$\begin{aligned}
\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom}} f} x) \in \operatorname{Argmin}\phi \quad&\Leftrightarrow\quad \xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom}} f} x) \in \operatorname{Argmin}\big(\phi/\gamma\big)\\
&\Leftrightarrow\quad \big(\mathrm{Id} - \operatorname{prox}_{\phi/\gamma}\big)\big(\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom}} f} x)\big) = 0\\
&\Leftrightarrow\quad \gamma^{-1}\operatorname{prox}_{\gamma\phi^*}\big(\xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom}} f} x)\big) = 0\\
&\Leftrightarrow\quad \operatorname{prox}_{\gamma\phi^*}\big(\xi^* + \gamma f(\operatorname{prox}_{\gamma(0\odot f)} x)\big) = 0\\
&\Leftrightarrow\quad \omega = 0.
\end{aligned} \tag{4.4}$$

- **Case 1:** $\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom}} f} x) \in \operatorname{Argmin}\phi$. Then (4.4) implies that $\omega = 0$ and therefore (4.2) yields

$$J_{\gamma\mathscr{K}_\Psi}(x,\xi^*) = \big(\operatorname{proj}_{\overline{\operatorname{dom}} f} x, 0\big). \tag{4.5}$$

- **Case 2:** $\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom}} f} x) \notin \operatorname{Argmin}\phi$. Then it follows from (4.4) that $\omega > 0$. Now define

$$T\colon \left]0,+\infty\right[\to \mathbb{R}\colon \tau \mapsto \tau - \operatorname{prox}_{\gamma\phi^*}\big(\xi^* + \gamma f(\operatorname{prox}_{\gamma\tau f} x)\big). \tag{4.6}$$

On the one hand, we deduce from [3, Lemma 3.27] and [8, Propositions 12.27 and 24.31] that T is continuous and strictly increasing. On the other hand, (4.1) and (2.8) imply that $\omega \in \operatorname{zer} T$, hence $\operatorname{zer} T = \{\omega\}$. Thus, ω can be computed by standard one-dimensional root-finding numerical schemes [43, Chapter 9]. In turn, (4.2) yields

$$J_{\gamma \mathscr{K}_\Psi}(x,\xi^*) = \big(\operatorname{prox}_{\gamma\omega f} x, \omega\big). \tag{4.7}$$

In the following examples, we compute the resolvent $J_{\gamma \mathscr{K}_\Psi}$ of Proposition 4.1 with the help of Remark 4.2.

Example 4.3 In Example 1.2, suppose that $\mathcal{X}$ is a real Hilbert space, let $\gamma \in \left]0,+\infty\right[$, let $(x,\xi^*) \in \mathcal{X}\times\mathbb{R}$, and let ω be the unique real number in $\left[0,+\infty\right[$ such that

$$\omega = \begin{cases} 0, & \text{if } \ \xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x) \leqslant 0;\\ \xi^* + \gamma f\big(\operatorname{prox}_{\gamma\omega f} x\big), & \text{if } \ \xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x) > 0. \end{cases} \tag{4.8}$$

Since $\phi^* = \iota_{\left[0,+\infty\right[}$ and $\operatorname{Argmin} \phi = \left]-\infty,0\right]$, (4.5) and (4.7) yield

$$J_{\gamma \mathscr{K}_\Psi}(x,\xi^*) = \begin{cases} \big(\operatorname{proj}_{\overline{\operatorname{dom} f}} x, 0\big), & \text{if } \ \xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x) \leqslant 0;\\ \big(\operatorname{prox}_{\gamma\omega f} x, \omega\big), & \text{if } \ \xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x) > 0. \end{cases} \tag{4.9}$$

Example 4.4 In Example 1.3, set $\alpha\colon \rho \mapsto \rho$, let $\gamma \in \left]0,+\infty\right[$, let $(x,\xi^*) \in \mathcal{X}\times\mathbb{R}$, and let ω be the unique real number in $\left[0,+\infty\right[$ such that

$$\omega = \begin{cases} 0, & \text{if } \ (\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x))/\rho \in \operatorname{Argmin} \theta;\\ \operatorname{prox}_{\gamma\rho\theta^*}\big(\xi^* + \gamma f(\operatorname{prox}_{\gamma\omega f} x)\big), & \text{if } \ (\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x))/\rho \notin \operatorname{Argmin} \theta. \end{cases} \tag{4.10}$$

Then, by [8, Proposition 13.23(ii)], $\phi^* = \rho\theta^*$, $\operatorname{Argmin} \phi = \rho \operatorname{Argmin} \theta$, and it follows from (4.5) and (4.7) that

$$J_{\gamma \mathscr{K}_\Psi}(x,\xi^*) = \begin{cases} \big(\operatorname{proj}_{\overline{\operatorname{dom} f}} x, 0\big), & \text{if } \ (\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x))/\rho \in \operatorname{Argmin} \theta;\\ \big(\operatorname{prox}_{\gamma\omega f} x, \omega\big), & \text{if } \ (\xi^*/\gamma + f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x))/\rho \notin \operatorname{Argmin} \theta. \end{cases} \tag{4.11}$$

Example 4.5 Consider the setting of Example 1.4, where we suppose that $\mathcal{X}$ is a real Hilbert space and that θ is not of the form $\theta = \iota_{\{0\}} + \nu$ for some $\nu \in \mathbb{R}$. Set $\rho = \sup \partial\theta(0)$, let $\gamma \in \left]0,+\infty\right[$, and let $(x,\xi^*) \in \mathcal{X}\times\mathbb{R}$. Then $\rho \in \left[0,+\infty\right]$, $\phi = \theta \circ d_{\left]-\infty,0\right]}$, $\operatorname{Argmin} \phi = \left]-\infty,0\right]$, and, since θ is even, [8, Example 13.26] yields $\gamma\phi^* = \gamma(\sigma_{\left]-\infty,0\right]} + \theta^* \circ |\cdot|) = \sigma_{\left]-\infty,0\right]} + \gamma\theta^* \circ |\cdot|$. In turn, since θ^* is not constant, it follows from [11, Proposition 2.2] that

$$\operatorname{prox}_{\gamma\phi^*} \xi^* = \begin{cases} 0, & \text{if } \ \xi^* \leqslant 0;\\ \xi^*, & \text{if } \ 0 < \xi^* \leqslant \rho;\\ \operatorname{prox}_{\gamma\theta^*} \xi^*, & \text{if } \ \xi^* > \rho. \end{cases} \tag{4.12}$$

Now let ω be the unique real number in $\left[0,+\infty\right[$ such that

$$\omega = \begin{cases} 0, & \text{if } \ \xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x) \leqslant 0;\\ \operatorname{prox}_{\gamma\phi^*}\big(\xi^* + \gamma f(\operatorname{prox}_{\gamma\omega f} x)\big), & \text{if } \ \xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x) > 0. \end{cases} \tag{4.13}$$

Then (4.5) and (4.7) yield

$$J_{\gamma \mathscr{K}_\Psi}(x,\xi^*) = \begin{cases} \big(\operatorname{proj}_{\overline{\operatorname{dom} f}} x, 0\big), & \text{if } \ \xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x) \leqslant 0;\\ \big(\operatorname{prox}_{\gamma\omega f} x, \omega\big), & \text{if } \ \xi^* + \gamma f(\operatorname{proj}_{\overline{\operatorname{dom} f}} x) > 0. \end{cases} \tag{4.14}$$

We now turn to the computation of the resolvents of $\boldsymbol{M}$, $\boldsymbol{S}$, and $\boldsymbol{G}$ in (3.30).

Proposition 4.6 *In Proposition 3.6, suppose that $\mathcal{X}$ and $\mathcal{Y}$ are real Hilbert spaces identified with their duals. Let $x \in \mathcal{X}$, $\xi^* \in \mathbb{R}$, $y^* \in \mathcal{Y}$, and $\gamma \in \,]0,+\infty[$. Then the following hold:*

(i) $J_{\gamma \boldsymbol{M}}(x,\xi^*,y^*) = (J_{\gamma\mathscr{K}_\Psi}(x,\xi^*), \operatorname{prox}_{\gamma g^*} y^*)$.

(ii) $J_{\gamma \boldsymbol{S}}(x,\xi^*,y^*) = ((\mathrm{Id} + \gamma^2 L^*L)^{-1}(x-\gamma L^*y^*), \xi^*, (\mathrm{Id} + \gamma^2 LL^*)^{-1}(y^*+\gamma Lx))$.

(iii) $J_{\gamma \boldsymbol{G}}(x,\xi^*,y^*) = \big(\operatorname{prox}_{\gamma h} x, \xi^*, \operatorname{prox}_{\gamma \ell^*} y^*\big)$.

Proof. (i): This follows from Proposition 3.6(ii), Lemma 2.11(ii), and [8, Proposition 23.18].

(ii): This follows from [8, Example 23.5].

(iii): This follows from [8, Proposition 23.18]. ☐

4.2 Algorithms

As an application of Theorem 3.3(iv), Proposition 3.6, and Proposition 4.6, we now design algorithms for solving the primal problem (1.2) and its dual (3.11). Several options can be considered to find a zero of $\mathscr{K}_F = \boldsymbol{M}+\boldsymbol{S}+\boldsymbol{G}$ depending on the assumptions on the constituents of the problem. In the following approach, we adopt a strategy based on the forward-backward-half forward splitting method of [14].

Notation 4.7 In the setting of Proposition 4.1, we denote by $\varpi(\gamma,x,\xi^*)$ the unique real number $\omega \in [0,+\infty[$ such that $\omega = \operatorname{prox}_{\gamma\phi^*}(\xi^* + \gamma f(\operatorname{prox}_{\gamma(\omega\odot f)} x))$.

Proposition 4.8 *Consider the setting of Problem 1.1, let $\mathscr{D}$ be the set of solutions to the dual problem (3.11), and define ϖ as in Notation 4.7. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are real Hilbert spaces, that $\mathscr{P} \neq \varnothing$, that h and ℓ^* are differentiable, and that ∇h and $\nabla \ell^*$ are δ- and ϑ-Lipschitzian, respectively, for some $(\delta,\vartheta) \in \,]0,+\infty[^2$. Let $(x_0,\xi_0^*,y_0^*) \in \mathcal{X}\times\mathbb{R}\times\mathcal{Y}$, let $\chi = 4\max\{\delta,\vartheta\}/(1+\sqrt{1+16\max\{\delta,\vartheta\}^2\|L\|^2})$, let $\varepsilon \in \,]0,\chi/2[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon,\chi-\varepsilon]$, and iterate*

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
z_n = x_n - \gamma_n\big(L^*y_n^* + \nabla h(x_n)\big)\\
z_n^* = y_n^* + \gamma_n\big(Lx_n - \nabla\ell^*(y_n^*)\big)\\
\xi_{n+1}^* = \varpi(\gamma_n, z_n, \xi_n^*)\\
p_n = \operatorname{prox}_{\gamma_n(\xi_{n+1}^*\odot f)} z_n\\
p_n^* = \operatorname{prox}_{\gamma_n g^*} z_n^*\\
x_{n+1} = p_n + \gamma_n L^*(y_n^* - p_n^*)\\
y_{n+1}^* = p_n^* - \gamma_n L(x_n - p_n).
\end{array}
\right.
\end{array}
\tag{4.15}
$$

Then $x_n \rightharpoonup \overline{x}$, $y_n^* \rightharpoonup \overline{y^*}$, and $\xi_n^* \to \overline{\xi^*}$, where $\overline{x} \in \mathscr{P}$ and $(\overline{\xi^*},\overline{y^*}) \in \mathscr{D}$.

Proof. Set $\boldsymbol{\mathcal{X}} = \mathcal{X}\oplus\mathbb{R}\oplus\mathcal{Y}$, let $(\boldsymbol{M},\boldsymbol{S},\boldsymbol{G})$ be as in (3.30), and let $\mathscr{K}_F$ be as in (3.12). We first note that Theorem 3.3(v) asserts that $\operatorname{zer}\mathscr{K}_F \neq \varnothing$. Further, it follows from Proposition 3.6(ii)–(v) that $\boldsymbol{M}$ is maximally monotone, that $\boldsymbol{S}$ is monotone and $\|L\|$-Lipschitzian, and that $\boldsymbol{G}$ is $(\max\{\delta,\vartheta\})^{-1}$-cocoercive. Now set, for every $n\in\mathbb{N}$, $\boldsymbol{x}_n = (x_n,\xi_n^*,y_n^*)$, $\boldsymbol{z}_n = (z_n,\xi_n^*,z_n^*)$, $\boldsymbol{p}_n = (p_n,\xi_{n+1}^*,p_n^*)$, and

$\boldsymbol{q}_n = (q_n, \xi^*_{n+1}, q^*_n)$. Then, in view of (3.30), Proposition 4.1, and Proposition 4.6(i), we can express (4.15) as

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \boldsymbol{z}_n = \boldsymbol{x}_n - \gamma_n(\boldsymbol{S}\boldsymbol{x}_n + \boldsymbol{G}\boldsymbol{x}_n) \\ \boldsymbol{p}_n = J_{\gamma_n \boldsymbol{M}}\, \boldsymbol{z}_n \\ \boldsymbol{x}_{n+1} = \boldsymbol{p}_n + \gamma_n \boldsymbol{S}(\boldsymbol{x}_n - \boldsymbol{p}_n). \end{array}\right.\end{array} \tag{4.16}$$

In turn, we derive from [14, Theorem 2.3.1] that $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point $\overline{\boldsymbol{x}} = (\overline{x}, \overline{\xi^*}, \overline{y^*}) \in \operatorname{zer} \mathscr{K}_F$. In view of Theorem 3.3(iv), the proof is complete. ☐

Remark 4.9 In Proposition 4.8, we have reduced solving the primal problem (1.2) and its dual (3.11) to finding a zero of the sum of a maximally monotone operator, a monotone Lipschitzian operator, and a cocoercive operator via the forward-backward-half forward splitting method [14]. Let us add a few comments.

(i) Suppose that $\ell^* = 0$ and $h = 0$. Then (4.16) is a nonlinear composite extension of the monotone+skew algorithm [12, Theorem 3.1]; the latter is obtained with $\phi\colon \xi \to \xi$. Another method tailored to inclusions involving the sum of a maximally monotone operator and a monotone Lipschitzian operator is that of [15, Corollary 5.2], which can also incorporate inertial effects. Another advantage of this framework is that it features, through [15, Theorem 4.8], a strongly convergent variant which does not require any additional assumptions on the operators. Other pertinent algorithms for this problem are found in [20, 38, 54].

(ii) Suppose that $g = 0$, $L = 0$, and $\ell^* = 0$. Then (4.16) is a nonlinear composite extension of of the forward-backward algorithm (see, e.g., [8, Corollary 28.9]); the latter is recovered with $\phi\colon \xi \to \xi$.

(iii) In the case when $h = 0$ and $\ell = \iota_{\{0\}}$, a zero of $\mathscr{K}_F = \boldsymbol{M} + \boldsymbol{S}$ can also be found by splitting methods which employ the resolvent of $\boldsymbol{S}$ given in Proposition 4.6(ii) and do not specifically exploit its Lipschitz continuity. See for instance [12, Remark 2.9].

A noteworthy special case of Problem 1.1 is when $\mathcal{Y} = \mathcal{X}$, $L = \mathrm{Id}$, $h = 0$, and $\ell = \iota_{\{0\}}$. In this setting, Proposition 3.2 asserts that Problem 1.1 can be solved by finding a zero of the sum of the monotone operators $\mathscr{K}_\Psi$ and $\boldsymbol{B}$ defined in (3.3) and (3.6), respectively. We can for instance use the Douglas–Rachford algorithm [8, Theorem 26.11] for this task, which leads to the following implementation.

Proposition 4.10 *Let $\mathcal{X}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{X})$, let $g \in \Gamma_0(\mathcal{X})$, let $\phi \in \Gamma_0(\mathbb{R})$ be increasing and not constant. Consider the minimization problem*

$$\underset{x\in\mathcal{X}}{\text{minimize}}\;\; (\phi \circ f)(x) + g(x), \tag{4.17}$$

under the assumption that solutions exist. Let $z_0 \in \mathcal{X}$, let $\zeta_0^* \in \mathbb{R}$, let $\gamma \in \left]0, +\infty\right[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, 2\right[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, define ϖ as in Notation 4.7, and iterate

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \xi^*_n = \varpi(\gamma, z_n, \zeta^*_n) \\ x_n = \operatorname{prox}_{\gamma(\xi^*_n \odot f)} z_n \\ z_{n+1} = z_n + \lambda_n\big(\operatorname{prox}_{\gamma g}(2x_n - z_n) - x_n\big) \\ \zeta^*_{n+1} = \zeta^*_n + \lambda_n(\xi^*_n - \zeta^*_n). \end{array}\right.\end{array} \tag{4.18}$$

Then $x_n \rightharpoonup \overline{x}$, where $\overline{x}$ is a solution to (4.17).

Proof. We first observe that (4.17) is the special case of Problem 1.1 corresponding to $\mathcal{Y} = \mathcal{X}$, $L = \mathrm{Id}$, $h = 0$, and $\ell = \iota_{\{0\}}$. It follows from Proposition 3.2(ii) that $\operatorname{zer}(\mathscr{K}_\Psi + \boldsymbol{B}) \neq \varnothing$, where $\mathscr{K}_\Psi$ is defined in (3.3) and $\boldsymbol{B}\colon (x, \xi^*) \mapsto \partial g(x) \times \{0\}$ follows from (3.6). Therefore, by [8, Proposition 23.18], $J_{\gamma \boldsymbol{B}}\colon (x, \xi^*) \mapsto (\operatorname{prox}_{\gamma g} x, \xi^*)$. Now set, for every $n \in \mathbb{N}$, $\boldsymbol{x}_n = (x_n, \xi_n^*) \in \mathcal{X} \oplus \mathbb{R}$ and $\boldsymbol{z}_n = (z_n, \zeta_n^*) \in \mathcal{X} \oplus \mathbb{R}$. Then it follows from Proposition 4.1 that (4.18) can be written as

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \boldsymbol{x}_n = J_{\gamma \mathscr{K}_\Psi} \boldsymbol{z}_n \\ \boldsymbol{z}_{n+1} = \boldsymbol{z}_n + \lambda_n \big(J_{\gamma \boldsymbol{B}}(2\boldsymbol{x}_n - \boldsymbol{z}_n) - \boldsymbol{x}_n \big). \end{array} \right. \tag{4.19}$$

In turn, [8, Theorem 26.11] yields $\boldsymbol{x}_n \rightharpoonup \overline{\boldsymbol{x}}$ for some $\overline{\boldsymbol{x}} \in \operatorname{zer}(\mathscr{K}_\Psi + \boldsymbol{B})$ and the result follows from Proposition 3.2(ii). ☐

Remark 4.11 In the case when $\phi\colon \xi \mapsto \xi$, $\phi^* = \iota_{\{1\}}$ and therefore $\varpi \equiv 1$. In this context, (4.15) reduces to the algorithm of [26, Theorem 3.1] and (4.18) reduces to the Douglas–Rachford algorithm [8, Theorem 26.11].

Remark 4.12

(i) As special cases, Examples 1.2–1.4 can be solved via Proposition 4.8 or Proposition 4.10, depending on the assumptions on L, ℓ, and h. In these settings, the function ϖ in (4.15) and (4.18) is computed in Examples 4.3–4.5.

(ii) In the particular case of Example 1.2, algorithms (4.15) and (4.18) activate the inequality constraint $f(x) \leqslant 0$ through the proximity operator of f, as described in Example 4.3. Note that general convex inequalities are hard to treat directly in the context of standard proximal methods since they involve the projection onto the 0-sublevel set of f, which is typically not explicit. For instance, in the case of (1.4), the projection onto the ℓ^p-ball is expensive to compute [24]. However, within our framework, (1.4) is solved by computing $\varpi(\gamma, x, \xi^*)$ via Example 4.3 as the unique real number in $\omega \in [0, +\infty[$ such that

$$\omega = \begin{cases} 0, & \text{if } \ \xi^*/\gamma + \|x\|_p^p \leqslant \eta^p; \\ \xi^* + \gamma\big(\|\operatorname{prox}_{\gamma\omega\|\cdot\|_p^p} x\|_p^p - \eta^p \big), & \text{if } \ \xi^*/\gamma + \|x\|_p^p > \eta^p, \end{cases} \tag{4.20}$$

while

$$\operatorname{prox}_{\gamma(\omega \odot \|\cdot\|_p^p)} x = \begin{cases} x, & \text{if } \ \xi^*/\gamma + \|x\|_p^p \leqslant \eta^p; \\ \big(\operatorname{prox}_{\gamma\omega|\cdot|^p} \xi_i\big)_{1 \leqslant i \leqslant N}, & \text{if } \ \xi^*/\gamma + \|x\|_p^p > \eta^p, \end{cases} \tag{4.21}$$

where $x = (\xi_i)_{1 \leqslant i \leqslant N}$ and $\operatorname{prox}_{\gamma\omega|\cdot|^p}$ is made explicit in [8, Example 24.38].

(iii) An alternative algorithmic setting that imposes no restriction on L, ℓ, and h in Problem 1.1 will be discussed in Remark 6.3.

5 Composite monotone inclusions

To study the general Problem 1.6, let us first go back to Problem 1.1 to motivate our approach. Define $\mathscr{K}_\Psi$ as in (3.3), $B = \partial g$ and set $D = \partial \ell$, $\boldsymbol{L}\colon \mathcal{X} \oplus \mathbb{R} \to \mathcal{Y}\colon (x, \xi^*) \mapsto Lx$, and $\boldsymbol{A}\colon \mathcal{X} \oplus \mathbb{R} \to 2^{\mathcal{X}^* \oplus \mathbb{R}}\colon (x, \xi^*) \mapsto (\partial h(x)) \times \{0\}$. Then it follows from Proposition 3.2(ii) that, if $\overline{\boldsymbol{z}} = (\overline{x}, \overline{\xi}^*) \in \mathcal{X} \oplus \mathbb{R}$ solves the monotone inclusion

$$\boldsymbol{0} \in \mathscr{K}_\Psi \overline{\boldsymbol{z}} + \big(\boldsymbol{L}^* \circ (B \,\square\, D) \circ \boldsymbol{L} \big) \overline{\boldsymbol{z}} + \boldsymbol{A} \overline{\boldsymbol{z}}, \tag{5.1}$$

then $\overline{x}$ solves Problem 1.1. Here, the Kuhn–Tucker operator $\mathscr{K}_\Psi$ is provided by the perturbation analysis of Proposition 3.1 and encapsulates the nonlinear composition in Problem 1.1. In turn, splitting algorithms are proposed in Section 4.2 to solve this inclusion.

We extend the above strategy by first defining in Section 5.1 adequate Kuhn–Tucker operators for Problem 1.6 through a perturbation analysis. In Section 5.2, we exploit this analysis to connect Problem 1.6 to a monotone inclusion problem that will allow us to solve it. In Section 5.3, we present an algorithm to solve the parent monotone inclusion problem and, thereby, Problem 1.6.

Our notational scheme is as follows.

Notation 5.1 In the setting of Problem 1.6, $\mathsf{x} = (\xi_i)_{i\in I}$ and $\mathsf{x}^* = (\xi_i^*)_{i\in I}$ are generic elements in $\mathbb{R}^I$, $\boldsymbol{z} = (x, \mathsf{x}) = (x, (\xi_i)_{i\in I})$ is a generic element in $\boldsymbol{\mathcal{Z}} = \mathcal{X} \oplus \mathbb{R}^I$, and $\boldsymbol{z}^* = (x^*, \mathsf{x}^*) = (x^*, (\xi_i^*)_{i\in I})$ is a generic element in $\boldsymbol{\mathcal{Z}}^* = \mathcal{X}^* \oplus \mathbb{R}^I$.

5.1 Perturbation analysis and Kuhn–Tucker operators

We introduce a perturbation analysis for the nonlinear compositions in Problem 1.6.

Proposition 5.2 *Consider the setting of Problem 1.6 and Notation 5.1. Fix $i \in I$, let*

$$\begin{aligned}&\Psi_i\colon \mathcal{X} \times \mathbb{R}^I \to \left]-\infty, +\infty\right]\\ &\big(x, (\xi_j)_{j\in I}\big) \mapsto \begin{cases} \phi_i\big(f_i(x) + \xi_i\big), & \text{if } f_i(x) + \xi_i \in \operatorname{dom}\phi_i \text{ and } (\forall j \in I \smallsetminus \{i\})\;\; \xi_j = 0;\\ +\infty, & \text{if } f_i(x) + \xi_i \notin \operatorname{dom}\phi_i \text{ or } (\exists\, j \in I \smallsetminus \{i\})\;\; \xi_j \neq 0\end{cases}\end{aligned} \tag{5.2}$$

be a perturbation of $\phi_i \circ f_i$, let $\mathscr{L}_{\Psi_i}$ be the associated Lagrangian, and let $\mathscr{K}_{\Psi_i}$ be the associated Kuhn–Tucker operator. Then the following hold:

(i) $\Psi_i \in \Gamma_0(\mathcal{X} \oplus \mathbb{R}^I)$.

(ii) *We have*

$$\mathscr{L}_{\Psi_i}\colon \boldsymbol{\mathcal{Z}} \to [-\infty, +\infty] : \boldsymbol{z} \mapsto \begin{cases} +\infty, & \text{if } x \notin \operatorname{dom} f_i;\\ (\xi_i^* \odot f_i)(x) - \phi_i^*(\xi_i^*), & \text{if } x \in \operatorname{dom} f_i \text{ and } \xi_i^* \in \operatorname{dom}\phi_i^*;\\ -\infty, & \text{if } x \in \operatorname{dom} f_i \text{ and } \xi_i^* \notin \operatorname{dom}\phi_i^*.\end{cases} \tag{5.3}$$

(iii) *Let $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$ and $\boldsymbol{z}^* \in \boldsymbol{\mathcal{Z}}^*$. Then*

$$\boldsymbol{z}^* \in \mathscr{K}_{\Psi_i}\boldsymbol{z} \quad\Leftrightarrow\quad \begin{cases} x \in \operatorname{dom} f_i \text{ and } \xi_i^* \in \operatorname{dom}\phi_i^*\\ x^* \in \partial(\xi_i^* \odot f_i)(x)\\ \xi_i \in \partial\phi_i^*(\xi_i^*) - f_i(x)\\ (\forall j \in I \smallsetminus \{i\})\; \xi_j = 0.\end{cases} \tag{5.4}$$

(iv) *$\mathscr{K}_{\Psi_i}$ is maximally monotone.*

Proof. (i): Set $\boldsymbol{f}_i\colon \mathcal{X} \oplus \mathbb{R} \to \left]-\infty, +\infty\right] : (x, \xi) \mapsto f_i(x) + \xi$. We derive from (5.2) and (1.1) that

$$\Psi_i\colon \big(x, (\xi_j)_{j\in I}\big) \mapsto \big(\phi_i \circ \boldsymbol{f}_i\big)(x, \xi_i) + \sum_{j\in I\smallsetminus\{i\}} \iota_{\{0\}}(\xi_j). \tag{5.5}$$

Since $f_i \in \Gamma_0(\mathcal{X})$, we have $\boldsymbol{f}_i \in \Gamma_0(\mathcal{X} \oplus \mathbb{R})$. Moreover, since $\varnothing \neq (\operatorname{dom}\phi_i) \cap f_i(\operatorname{dom} f_i) \subset (\operatorname{dom}\phi_i) \cap \boldsymbol{f}_i(\operatorname{dom}\boldsymbol{f}_i)$, Proposition 2.9(ii) asserts that $\phi_i \circ \boldsymbol{f}_i \in \Gamma_0(\mathcal{X} \oplus \mathbb{R})$ and the claim follows.

(ii): Let $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. We deduce from (2.24) that

$$
\begin{aligned}
\mathscr{L}_{\Psi_i}(\boldsymbol{z}) &= \inf_{(\xi_j)_{j\in I}\in\mathbb{R}^I} \left(\Psi_i\big(x,(\xi_j)_{j\in I}\big) - \sum_{j\in I} \xi_j\xi_j^* \right) \\
&= \begin{cases} -\displaystyle\sup_{\xi_i\in(\operatorname{dom}\phi_i)-f_i(x)} \big(\xi_i\xi_i^* - \phi_i(f_i(x)+\xi_i)\big), & \text{if } x \in \operatorname{dom} f_i; \\ +\infty, & \text{if } x \notin \operatorname{dom} f_i \end{cases} \\
&= \begin{cases} \xi_i^* f_i(x) - \displaystyle\sup_{\xi_i+f_i(x)\in\operatorname{dom}\phi_i} \big(\xi_i + f_i(x)\big)\xi_i^* - \phi_i\big(f_i(x)+\xi_i\big), & \text{if } x \in \operatorname{dom} f_i; \\ +\infty, & \text{if } x \notin \operatorname{dom} f_i \end{cases} \\
&= \begin{cases} \xi_i^* f_i(x) - \phi_i^*(\xi_i^*), & \text{if } x \in \operatorname{dom} f_i; \\ +\infty, & \text{if } x \notin \operatorname{dom} f_i. \end{cases}
\end{aligned} \tag{5.6}
$$

The claim therefore follows from Lemma 2.8(ii) and (2.8).

(iii): Note that, if $x \notin \operatorname{dom} f_i$, (ii) yields $\mathscr{L}_{\Psi_i}(\boldsymbol{z}) = +\infty$ and, hence, $\mathscr{K}_{\Psi_i}(\boldsymbol{z}) = \varnothing$ in view of (2.26). Similarly, if $\xi_i^* \notin \operatorname{dom}\phi_i^*$, then $-\mathscr{L}_{\Psi_i}(\boldsymbol{z}) = +\infty$, which yields $\mathscr{K}_{\Psi_i}(\boldsymbol{z}) = \varnothing$. Now suppose that $x \in \operatorname{dom} f_i$ and $\xi_i^* \in \operatorname{dom}\phi_i^*$. Arguing along the same lines as in the proof of Proposition 3.1(iii), we deduce that $\partial(\mathscr{L}_{\Psi_i}(\cdot,(\xi_j^*)_{j\in I})) = \partial(\xi_i^* \odot f_i)$ and that $(\xi_j)_{j\in I} \in \partial\big(-\mathscr{L}_{\Psi_i}(x,\cdot)\big)(\xi_j^*)_{j\in I}$ if and only if $\xi_i \in \partial\phi_i^*(\xi_i^*) - f_i(x)$ and, for every $j \in I \smallsetminus \{i\}$, $\xi_j = 0$. Altogether, (2.26) yields (5.4).

(iv): This follows from (i) and Lemma 2.11(ii). ☐

5.2 Monotone inclusion formulation

We associate Problem 1.6 with a monotone inclusion problem.

Proposition 5.3 *Consider the setting of Problem 1.6 and Notation 5.1. Define the operators $(\mathscr{K}_{\Psi_i})_{i\in I}$ as in (5.4), and set*

$$
\begin{cases} \boldsymbol{A}\colon \boldsymbol{\mathcal{Z}} \to 2^{\boldsymbol{\mathcal{Z}}^*} \colon \boldsymbol{z} \mapsto Ax \times \{0\} \\ \boldsymbol{C}\colon \boldsymbol{\mathcal{Z}} \to \boldsymbol{\mathcal{Z}}^* \colon \boldsymbol{z} \mapsto (Cx, 0) \\ \boldsymbol{Q}\colon \boldsymbol{\mathcal{Z}} \to \boldsymbol{\mathcal{Z}}^* \colon \boldsymbol{z} \mapsto (Qx, 0) \\ (\forall k \in K)\ \ \boldsymbol{L}_k\colon \boldsymbol{\mathcal{Z}} \to \mathcal{Y}_k \colon \boldsymbol{z} \mapsto L_k x. \end{cases} \tag{5.7}
$$

Consider the inclusion problem

$$
\text{find } \ \boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \ \text{ such that } \ \boldsymbol{0} \in \sum_{i\in I} \mathscr{K}_{\Psi_i}\boldsymbol{z} + \sum_{k\in K} \big(\boldsymbol{L}_k^* \circ (B_k \,\square\, D_k) \circ \boldsymbol{L}_k\big)\boldsymbol{z} + \boldsymbol{A}\boldsymbol{z} + \boldsymbol{C}\boldsymbol{z} + \boldsymbol{Q}\boldsymbol{z}. \tag{5.8}
$$

Then the following hold:

(i) *Suppose that $\overline{\boldsymbol{z}} = (\overline{x}, (\overline{\xi_i^*})_{i\in I}) \in \boldsymbol{\mathcal{Z}}$ solves (5.8). Then $\overline{x}$ solves Problem 1.6.*

(ii) *Suppose that $\overline{x} \in \mathcal{X}$ solves Problem 1.6 and that, for every $i \in I$, $(\operatorname{int}\operatorname{dom}\phi_i) \cap f_i(\operatorname{dom} f_i) \neq \varnothing$. Then there exists $(\overline{\xi_i^*})_{i\in I} \in \mathbb{R}^I$ such that $(\overline{x}, (\overline{\xi_i^*})_{i\in I})$ solves (5.8).*

Proof. In view of (5.4) and (5.7), for every $\overline{\boldsymbol{z}} = (\overline{x}, (\overline{\xi_i^*})_{i\in I}) \in \boldsymbol{\mathcal{Z}}$,

$$
\begin{aligned}
\boldsymbol{0} \;&\in\; \sum_{i\in I}\mathscr{K}_{\Psi_i}\overline{\boldsymbol{z}} + \sum_{k\in K}\big(\boldsymbol{L}_k^*\circ(B_k\,\square\,D_k)\circ\boldsymbol{L}_k\big)\overline{\boldsymbol{z}} + \boldsymbol{A}\overline{\boldsymbol{z}} + \boldsymbol{C}\overline{\boldsymbol{z}} + \boldsymbol{Q}\overline{\boldsymbol{z}}\\
&\Leftrightarrow\; \begin{cases} 0 \in \displaystyle\sum_{i\in I}\partial(\overline{\xi_i^*}\odot f_i)(\overline{x}) + \sum_{k\in K}\big(L_k^*\circ(B_k\,\square\,D_k)\circ L_k\big)\overline{x} + A\overline{x} + C\overline{x} + Q\overline{x}\\ (\forall i\in I)\quad 0 \in \partial\phi_i^*(\overline{\xi_i^*}) - f_i(\overline{x}),\end{cases}\\
&\Leftrightarrow\; \begin{cases} 0 \in \displaystyle\sum_{i\in I}\partial(\overline{\xi_i^*}\odot f_i)(\overline{x}) + \sum_{k\in K}\big(L_k^*\circ(B_k\,\square\,D_k)\circ L_k\big)\overline{x} + A\overline{x} + C\overline{x} + Q\overline{x}\\ (\forall i\in I)\quad \overline{\xi_i^*} \in \partial\phi_i\big(f_i(\overline{x})\big).\end{cases}
\end{aligned}
\tag{5.9}
$$

(i): This follows from (5.9) and Proposition 2.9(iv).

(ii): This follows from Proposition 2.9(v) and (5.9). ◻

5.3 Block-iterative algorithm

We provide a flexible splitting algorithm for solving Problem 1.6, in which all the operators and functions present in (1.11) are activated separately. The proposed algorithm allows for block-iterative implementations in the sense that, at a given iteration, only a subgroup of functions and operators needs to be employed, as opposed to all of them as in classical methods. This feature is particularly important in large-scale applications, where it is either impossible or inefficient to evaluate each the operators at each iteration. This algorithm is an offspring of that of [16], where one will find more details on its construction. In a nutshell, its principle is to recast the parent inclusion problem (5.8) as that of finding a zero of a saddle operator that acts on a bigger space. This task is achieved by successive projections onto half-spaces containing the set of zeros of the saddle operator. At a given iteration, the construction of such a half-space involves the resolvents of the selected operators.

The following vector version of Notation 4.7 will be needed.

Notation 5.4 Let $\gamma \in \left]0,+\infty\right[$, let $x \in \mathcal{X}$, and let $\mathsf{x} \in \mathbb{R}^I$. For every $i \in I$, let ϖ_i be the function defined as in Notation 4.7 with respect to (ϕ_i, f_i), let $\mathsf{e}_i \in \mathbb{R}^I$ be the ith canonical vector, let $[\mathsf{x}]_i$ be the ith coordinate of x, and set $\mathsf{w}_i(\gamma, x, \mathsf{x}) = \mathsf{x} + (\varpi_i(\gamma, x, [\mathsf{x}]_i) - [\mathsf{x}]_i)\mathsf{e}_i$.

Theorem 5.5 *Consider the setting of Problem 1.6 and of Notations 2.1, 5.1, and 5.4. Suppose that $\mathcal{X}$ and $(\mathcal{Y}_k)_{k\in K}$ are real Hilbert spaces with scalar products denoted by $\langle\cdot\mid\cdot\rangle$ and that Problem 1.6 admits at least one solution. In addition, let $\sigma \in \left]1/(4\beta),+\infty\right[$, let $\varepsilon \in \left]0,\min\{1,1/(\chi+\sigma)\}\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ be sequences in $[\varepsilon, 1/(\chi+\sigma)]$ and $[\varepsilon, 2-\varepsilon]$, respectively, and, for every $j \in I\cup K$, let $(\mu_{j,n})_{n\in\mathbb{N}}$, $(\nu_{j,n})_{n\in\mathbb{N}}$, and $(\sigma_{j,n})_{n\in\mathbb{N}}$ be sequences in $[\varepsilon, 1/\sigma]$, $[\varepsilon, 1/\sigma]$, and $[\varepsilon, 1/\varepsilon]$, respectively. Moreover, let $x_0 \in \mathcal{X}$, let $\mathsf{x}_0 \in \mathbb{R}^I$, for every $i \in I$, let $\{y_{i,0}, u_{i,0}, v_{i,0}^*\} \subset \mathcal{X}$ and $\{\mathsf{y}_{i,0}, \mathsf{u}_{i,0}, \mathsf{v}_{i,0}^*\} \subset \mathbb{R}^I$, and, for every $k \in K$, let $\{y_{k,0}, u_{k,0}, v_{k,0}^*\} \subset \mathcal{Y}_k$. Further, let $(P_n)_{n\in\mathbb{N}}$ be nonempty subsets of $I\cup K$ such that*

$$
P_0 = I\cup K \quad \textit{and} \quad (\exists\, M \in \mathbb{N})(\forall n \in \mathbb{N}) \quad \bigcup_{k=n}^{n+M} P_k = I\cup K. \tag{5.10}
$$

Iterate

for $n = 0, 1, \ldots$

$l_n^* = Qx_n + \sum_{i\in I} v_{i,n}^* + \sum_{k\in K} L_k^* v_{k,n}^*;\ \ell_n^* = \sum_{i\in I} \mathsf{v}_{i,n}^*;$

$a_n = J_{\gamma_n A}\big(x_n - \gamma_n(l_n^* + Cx_n)\big);\ \mathsf{a}_n = \mathsf{x}_n - \gamma_n \ell_n^*;$

$a_n^* = \gamma_n^{-1}(x_n - a_n) - l_n^* + Qa_n;\ \xi_n = \|a_n - x_n\|^2 + \|\mathsf{a}_n - \mathsf{x}_n\|^2;$

for every $i \in I \cap P_n$

$\mathsf{b}_{i,n} = \mathsf{w}_i(\mu_{i,n}, y_{i,n} + \mu_{i,n} v_{i,n}^*, \mathsf{y}_{i,n} + \mu_{i,n}\mathsf{v}_{i,n}^*);\ b_{i,n} = \operatorname{prox}_{\mu_{i,n}([\mathsf{b}_{i,n}]_i \odot f_i)}\big(y_{i,n} + \mu_{i,n} v_{i,n}^*\big);$

$e_{i,n} = b_{i,n} - a_n;\ \mathsf{e}_{i,n} = \mathsf{b}_{i,n} - \mathsf{a}_n;$

$e_{i,n}^* = \sigma_{i,n}(x_n - y_{i,n} - u_{i,n}) + v_{i,n}^*;\ \mathsf{e}_{i,n}^* = \sigma_{i,n}(\mathsf{x}_n - \mathsf{y}_{i,n} - \mathsf{u}_{i,n}) + \mathsf{v}_{i,n}^*;$

$q_{i,n}^* = \mu_{i,n}^{-1}(y_{i,n} - b_{i,n}) + v_{i,n}^* - e_{i,n}^*;\ \mathsf{q}_{i,n}^* = \mu_{i,n}^{-1}(\mathsf{y}_{i,n} - \mathsf{b}_{i,n}) + \mathsf{v}_{i,n}^* - \mathsf{e}_{i,n}^*;$

$t_{i,n}^* = \nu_{i,n}^{-1} u_{i,n} + v_{i,n}^* - e_{i,n}^*;\ \mathsf{t}_{i,n}^* = \nu_{i,n}^{-1}\mathsf{u}_{i,n} + \mathsf{v}_{i,n}^* - \mathsf{e}_{i,n}^*;$

$\eta_{i,n} = \|b_{i,n} - y_{i,n}\|^2 + \|\mathsf{b}_{i,n} - \mathsf{y}_{i,n}\|^2 + \|u_{i,n}\|^2 + \|\mathsf{u}_{i,n}\|^2;$

for every $k \in K \cap P_n$

$b_{k,n} = J_{\mu_{k,n} B_k}\big(y_{k,n} + \mu_{k,n} v_{k,n}^*\big);\ d_{k,n} = J_{\nu_{k,n} D_k}\big(u_{k,n} + \nu_{k,n} v_{k,n}^*\big);$

$e_{k,n} = b_{k,n} + d_{k,n} - L_k a_n;\ e_{k,n}^* = \sigma_{k,n}\big(L_k x_n - y_{k,n} - u_{k,n}\big) + v_{k,n}^*;$

$q_{k,n}^* = \mu_{k,n}^{-1}(y_{k,n} - b_{k,n}) + v_{k,n}^* - e_{k,n}^*;\ t_{k,n}^* = \nu_{k,n}^{-1}(u_{k,n} - d_{k,n}) + v_{k,n}^* - e_{k,n}^*;$

$\eta_{k,n} = \|b_{k,n} - y_{k,n}\|^2 + \|d_{k,n} - u_{k,n}\|^2;$

for every $i \in I \smallsetminus P_n$

$b_{i,n} = b_{i,n-1};\ \mathsf{b}_{i,n} = \mathsf{b}_{i,n-1};\ e_{i,n} = b_{i,n} - a_n;\ \mathsf{e}_{i,n} = \mathsf{b}_{i,n} - \mathsf{a}_n; e_{i,n}^* = e_{i,n-1}^*;$

$\mathsf{e}_{i,n}^* = \mathsf{e}_{i,n-1}^*;\ q_{i,n}^* = q_{i,n-1}^*;\ \mathsf{q}_{i,n}^* = \mathsf{q}_{i,n-1}^*;\ t_{i,n}^* = t_{i,n-1}^*;\ \mathsf{t}_{i,n}^* = \mathsf{t}_{i,n-1}^*;\ \eta_{i,n} = \eta_{i,n-1};$

for every $k \in K \smallsetminus P_n$

$b_{k,n} = b_{k,n-1};\ d_{k,n} = d_{k,n-1};\ e_{k,n} = b_{k,n} + d_{k,n} - L_k a_n;\ e_{k,n}^* = e_{k,n-1}^*;$

$q_{k,n}^* = q_{k,n-1}^*;\ t_{k,n}^* = t_{k,n-1}^*;\ \eta_{k,n} = \eta_{k,n-1};$

$$p_n^* = a_n^* + \sum_{i\in I} e_{i,n}^* + \sum_{k\in K} L_k^* e_{k,n}^*;\ \mathsf{p}_n^* = \sum_{i\in I} \mathsf{e}_{i,n}^*;$$

$$\begin{aligned}\Delta_n = {} & -(4\beta)^{-1}\big(\xi_n + \textstyle\sum_{j\in I\cup K} \eta_{j,n}\big) + \langle x_n - a_n \mid p_n^*\rangle + \langle \mathsf{x}_n - \mathsf{a}_n \mid \mathsf{p}_n^*\rangle \\ & + \textstyle\sum_{i\in I} \big(\langle y_{i,n} - b_{i,n} \mid q_{i,n}^*\rangle + \langle u_{i,n} \mid t_{i,n}^*\rangle + \langle e_{i,n} \mid v_{i,n}^* - e_{i,n}^*\rangle\big) \\ & + \textstyle\sum_{i\in I} \big(\langle \mathsf{y}_{i,n} - \mathsf{b}_{i,n} \mid \mathsf{q}_{i,n}^*\rangle + \langle \mathsf{u}_{i,n} \mid \mathsf{t}_{i,n}^*\rangle + \langle \mathsf{e}_{i,n} \mid \mathsf{v}_{i,n}^* - \mathsf{e}_{i,n}^*\rangle\big) \\ & + \textstyle\sum_{k\in K} \big(\langle y_{k,n} - b_{k,n} \mid q_{k,n}^*\rangle + \langle u_{k,n} - d_{k,n} \mid t_{k,n}^*\rangle + \langle e_{k,n} \mid v_{k,n}^* - e_{k,n}^*\rangle\big);\end{aligned}$$

if $\Delta_n > 0$

$$\alpha_n = \|p_n^*\|^2 + \|\mathsf{p}_n^*\|^2 + \sum_{j\in I\cup K} \|q_{j,n}^*\|^2 + \|t_{j,n}^*\|^2 + \|e_{j,n}\|^2 + \sum_{i\in I} \|\mathsf{q}_{i,n}^*\|^2 + \|\mathsf{t}_{i,n}^*\|^2 + \|\mathsf{e}_{i,n}\|^2;$$

$\theta_n = \lambda_n \Delta_n / \alpha_n;$

$x_{n+1} = x_n - \theta_n p_n^*;\ \mathsf{x}_{n+1} = \mathsf{x}_n - \theta_n \mathsf{p}_n^*;$

for every $j \in I \cup K$

$y_{j,n+1} = y_{j,n} - \theta_n q_{j,n}^*;\ u_{j,n+1} = u_{j,n} - \theta_n t_{j,n}^*;\ v_{j,n+1}^* = v_{j,n}^* - \theta_n e_{j,n};$

for every $i \in I$

$\mathsf{y}_{i,n+1} = \mathsf{y}_{i,n} - \theta_n \mathsf{q}_{i,n}^*;\ \mathsf{u}_{i,n+1} = \mathsf{u}_{i,n} - \theta_n \mathsf{t}_{i,n}^*;\ \mathsf{v}_{i,n+1}^* = \mathsf{v}_{i,n}^* - \theta_n \mathsf{e}_{i,n};$

else

$x_{n+1} = x_n;\ \mathsf{x}_{n+1} = \mathsf{x}_n;$

for every $j \in I \cup K$

$y_{j,n+1} = y_{j,n};\ u_{j,n+1} = u_{j,n};\ v_{j,n+1}^* = v_{j,n}^*;$

for every $i \in I$

$\mathsf{y}_{i,n+1} = \mathsf{y}_{i,n};\ \mathsf{u}_{i,n+1} = \mathsf{u}_{i,n};\ \mathsf{v}_{i,n+1}^* = \mathsf{v}_{i,n}^*.$

(5.11)

Then $x_n \rightharpoonup \overline{x}$, where $\overline{x}$ is a solution to Problem 1.6.

Proof. Set $P = I \cup K$, let $(\mathscr{K}_{\Psi_i})_{i\in I}$, $(\boldsymbol{L}_k)_{k\in K}$, $\boldsymbol{A}$, $\boldsymbol{C}$, and $\boldsymbol{Q}$ be the operators defined in (5.4) and (5.7), and, for every $j \in P$, define

$$\boldsymbol{B}_j = \begin{cases} \mathscr{K}_{\Psi_j}, & \text{if } j \in I;\\ B_j, & \text{if } j \in K, \end{cases} \quad \boldsymbol{D}_j = \begin{cases} N_{\{0\}}, & \text{if } j \in I;\\ D_j, & \text{if } j \in K, \end{cases} \quad \text{and} \quad \boldsymbol{M}_j = \begin{cases} \mathbf{Id}, & \text{if } j \in I;\\ \boldsymbol{L}_j, & \text{if } j \in K. \end{cases} \tag{5.12}$$

Then $\boldsymbol{A}$ is maximally monotone, $\boldsymbol{C}$ is β-cocoercive, $\boldsymbol{Q}$ is monotone and χ-Lipschitzian, and, for every $j \in P$, $\boldsymbol{B}_j$ and $\boldsymbol{D}_j$ are maximally monotone (see Proposition 5.2(iv)) and $\boldsymbol{M}_j$ is linear and bounded. It follows from Proposition 5.3(ii) and (5.12) that there exists $\overline{\boldsymbol{z}} = (\overline{x}, (\overline{\xi_i^*})_{i\in I}) \in \mathcal{X} \oplus \mathbb{R}^I$ such that

$$\mathbf{0} \in \sum_{j\in P} \big(\boldsymbol{M}_j^* \circ (\boldsymbol{B}_j \,\square\, \boldsymbol{D}_j) \circ \boldsymbol{M}_j\big)\overline{\boldsymbol{z}} + \boldsymbol{A}\overline{\boldsymbol{z}} + \boldsymbol{C}\overline{\boldsymbol{z}} + \boldsymbol{Q}\overline{\boldsymbol{z}}. \tag{5.13}$$

Moreover, we derive from (5.4) and Proposition 4.1 that

$$(\forall i \in I)(\forall \gamma \in \left]0,+\infty\right[) \quad J_{\gamma\mathscr{K}_{\Psi_i}} \colon (x, \mathsf{x}) \mapsto \big(\mathrm{prox}_{\gamma(\varpi_i(\gamma,x,[\mathsf{x}]_i)\odot f_i)} x, \mathsf{w}_i(\gamma, x, \mathsf{x})\big). \tag{5.14}$$

Now, for every $n \in \mathbb{N}$, set

$$\boldsymbol{z}_n = (x_n, \mathsf{x}_n),\ \boldsymbol{l}_n^* = (l_n^*, \ell_n^*),\ \boldsymbol{a}_n = (a_n, \mathsf{a}_n),\ \boldsymbol{a}_n^* = (a_n^*, 0),\ \boldsymbol{p}_n^* = (p_n^*, \mathsf{p}_n^*), \tag{5.15}$$

for every $j \in I$, set

$$\begin{cases} \boldsymbol{b}_{j,n} = (b_{j,n}, \mathsf{b}_{j,n}),\ \boldsymbol{e}_{j,n}^* = (e_{j,n}^*, \mathsf{e}_{j,n}^*),\ \boldsymbol{e}_{j,n} = (e_{j,n}, \mathsf{e}_{j,n}),\ \boldsymbol{q}_{j,n}^* = (q_{j,n}^*, \mathsf{q}_{j,n}^*),\ \boldsymbol{t}_{j,n}^* = (t_{j,n}^*, \mathsf{t}_{j,n}^*),\\ \boldsymbol{y}_{j,n} = (y_{j,n}, \mathsf{y}_{j,n}),\ \boldsymbol{u}_{j,n} = (u_{j,n}, \mathsf{u}_{j,n}),\ \boldsymbol{v}_{j,n}^* = (v_{j,n}^*, \mathsf{v}_{j,n}^*),\ \boldsymbol{d}_{j,n} = (0,0), \end{cases} \tag{5.16}$$

and, for every $j \in K$, set

$$\begin{cases} \boldsymbol{b}_{j,n} = b_{j,n},\ \boldsymbol{e}_{j,n}^* = e_{j,n}^*,\ \boldsymbol{e}_{j,n} = e_{j,n},\ \boldsymbol{q}_{j,n}^* = q_{j,n}^*,\ \boldsymbol{t}_{j,n}^* = t_{j,n}^*,\\ \boldsymbol{y}_{j,n} = y_{j,n},\ \boldsymbol{u}_{j,n} = u_{j,n},\ \boldsymbol{v}_{j,n}^* = v_{j,n}^*,\ \boldsymbol{d}_{j,n} = d_{j,n}. \end{cases} \tag{5.17}$$

Using (5.12), (5.14), (5.15), (5.16), (5.17), and elementary manipulations, we reduce (5.11) to

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots\\
\left\lfloor \begin{array}{l}
\boldsymbol{l}_n^* = \boldsymbol{Q}\boldsymbol{z}_n + \sum_{j\in P} \boldsymbol{M}_j^* \boldsymbol{v}_{j,n}^*;\\
\boldsymbol{a}_n = J_{\gamma_n \boldsymbol{A}}\big(\boldsymbol{z}_n - \gamma_n(\boldsymbol{l}_n^* + \boldsymbol{C}\boldsymbol{z}_n)\big);\\
\boldsymbol{a}_n^* = \gamma_n^{-1}(\boldsymbol{z}_n - \boldsymbol{a}_n) - \boldsymbol{l}_n^* + \boldsymbol{Q}\boldsymbol{a}_n;\\
\xi_n = \|\boldsymbol{a}_n - \boldsymbol{z}_n\|^2;\\
\text{for every } j \in P_n\\
\left\lfloor \begin{array}{l}
\boldsymbol{b}_{j,n} = J_{\mu_{j,n}\boldsymbol{B}_j}\big(\boldsymbol{y}_{j,n} + \mu_{j,n}\boldsymbol{v}_{j,n}^*\big);\\
\boldsymbol{d}_{j,n} = J_{\nu_{j,n}\boldsymbol{D}_j}\big(\boldsymbol{u}_{j,n} + \nu_{j,n}\boldsymbol{v}_{j,n}^*\big);\\
\boldsymbol{e}_{j,n} = \boldsymbol{b}_{j,n} + \boldsymbol{d}_{j,n} - \boldsymbol{M}_j\boldsymbol{a}_n;\\
\boldsymbol{e}_{j,n}^* = \sigma_{j,n}\big(\boldsymbol{M}_j\boldsymbol{z}_n - \boldsymbol{y}_{j,n} - \boldsymbol{u}_{j,n}\big) + \boldsymbol{v}_{j,n}^*;\\
\boldsymbol{q}_{j,n}^* = \mu_{j,n}^{-1}(\boldsymbol{y}_{j,n} - \boldsymbol{b}_{j,n}) + \boldsymbol{v}_{j,n}^* - \boldsymbol{e}_{j,n}^*;\\
\boldsymbol{t}_{j,n}^* = \nu_{j,n}^{-1}(\boldsymbol{u}_{j,n} - \boldsymbol{d}_{j,n}) + \boldsymbol{v}_{j,n}^* - \boldsymbol{e}_{j,n}^*;\\
\eta_{j,n} = \|\boldsymbol{b}_{j,n} - \boldsymbol{y}_{j,n}\|^2 + \|\boldsymbol{d}_{j,n} - \boldsymbol{u}_{j,n}\|^2;
\end{array}\right.
\end{array}\right.
\end{array} \tag{5.18}$$

$$
\begin{array}{l}
\text{for every } j \in P \smallsetminus P_n \\
\quad \boldsymbol{b}_{j,n} = \boldsymbol{b}_{j,n-1};\ \boldsymbol{d}_{j,n} = \boldsymbol{d}_{j,n-1};\ \boldsymbol{e}^*_{j,n} = \boldsymbol{e}^*_{j,n-1};\ \boldsymbol{q}^*_{j,n} = \boldsymbol{q}^*_{j,n-1};\ \boldsymbol{t}^*_{j,n} = \boldsymbol{t}^*_{j,n-1}; \\
\quad \boldsymbol{e}_{j,n} = \boldsymbol{b}_{j,n} + \boldsymbol{d}_{j,n} - \boldsymbol{M}_j \boldsymbol{a}_n;\ \eta_{j,n} = \eta_{j,n-1}; \\
\boldsymbol{p}^*_n = \boldsymbol{a}^*_n + \sum_{j\in P} \boldsymbol{M}^*_j \boldsymbol{e}^*_{j,n}; \\
\Delta_n = -(4\beta)^{-1}\big(\xi_n + \sum_{j\in P}\eta_{j,n}\big) + \langle \boldsymbol{z}_n - \boldsymbol{a}_n \mid \boldsymbol{p}^*_n\rangle \\
\qquad + \sum_{j\in P}\big(\langle \boldsymbol{y}_{j,n} - \boldsymbol{b}_{j,n} \mid \boldsymbol{q}^*_{j,n}\rangle + \langle \boldsymbol{u}_{j,n} - \boldsymbol{d}_{j,n} \mid \boldsymbol{t}^*_{j,n}\rangle + \langle \boldsymbol{e}_{j,n} \mid \boldsymbol{v}^*_{j,n} - \boldsymbol{e}^*_{j,n}\rangle\big); \\
\text{if } \Delta_n > 0 \\
\quad \theta_n = \lambda_n\Delta_n/\big(\|\boldsymbol{p}^*_n\|^2 + \sum_{j\in P}\big(\|\boldsymbol{q}^*_{j,n}\|^2 + \|\boldsymbol{t}^*_{j,n}\|^2 + \|\boldsymbol{e}_{j,n}\|^2\big)\big); \\
\quad \boldsymbol{z}_{n+1} = \boldsymbol{z}_n - \theta_n \boldsymbol{p}^*_n; \\
\quad \text{for every } j \in P \\
\quad\quad \boldsymbol{y}_{j,n+1} = \boldsymbol{y}_{j,n} - \theta_n \boldsymbol{q}^*_{j,n};\ \boldsymbol{u}_{j,n+1} = \boldsymbol{u}_{j,n} - \theta_n \boldsymbol{t}^*_{j,n};\ \boldsymbol{v}^*_{j,n+1} = \boldsymbol{v}^*_{j,n} - \theta_n \boldsymbol{e}_{j,n}; \\
\text{else} \\
\quad \boldsymbol{z}_{n+1} = \boldsymbol{z}_n; \\
\quad \text{for every } j \in P \\
\quad\quad \boldsymbol{y}_{j,n+1} = \boldsymbol{y}_{j,n};\ \boldsymbol{u}_{j,n+1} = \boldsymbol{u}_{j,n};\ \boldsymbol{v}^*_{j,n+1} = \boldsymbol{v}^*_{j,n}.
\end{array}
$$

Since (5.13) corresponds to a particular instance of [16, Problem 1] and (5.18) is the application of [16, Algorithm 1] to solve (5.13), the result follows from [16, Theorem 1(iv)]. ☐

Remark 5.6

(i) The weakly convergent algorithm (5.11) is based on [16, Algorithm 1]. We can also derive from [16, Theorem 2(iv)] a strongly convergent variant of it which does not require additional assumptions on Problem 1.6.

(ii) Although this point is omitted for simplicity, [16, Algorithm 1] and [16, Theorem 2(iv)] allow for asynchronous implementations, in the sense that any iteration can incorporate the result of calculations initiated at earlier iterations. In turn, (5.11) (and its strongly convergent counterpart) can also be executed asynchronously.

6 Applications

We describe a few applications of our framework to variational problems.

Example 6.1 In Problem 1.6, suppose that, for every $k \in K$, $B_k = N_{E_k}$ and $D_k = N_{F_k}$, where E_k and F_k are nonempty closed convex subsets of $\mathcal{Y}_k$ such that $0 \in \operatorname{sri}(\operatorname{bar} E_k - \operatorname{bar} F_k)$. Define

$$G = \big\{x \in \mathcal{X} \mid (\forall k \in K)\ L_k x \in E_k + F_k\big\}. \tag{6.1}$$

Note that, for every $k \in K$, $0 \in \operatorname{sri}(\operatorname{bar} E_k - \operatorname{bar} F_k) = \operatorname{sri}(\operatorname{dom} \iota^*_{E_k} - \operatorname{dom} \iota^*_{F_k})$ and, therefore, Lemma 2.7 and (6.1) imply that

$$\begin{aligned}\sum_{k\in K} L^*_k \circ (N_{E_k} \,\square\, N_{F_k}) \circ L_k &= \sum_{k\in K} L^*_k \circ (\partial\iota_{E_k} \,\square\, \partial\iota_{F_k}) \circ L_k \\ &= \sum_{k\in K} L^*_k \circ \partial(\iota_{E_k} \,\square\, \iota_{F_k}) \circ L_k\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k\in K} L_k^* \circ \partial(\iota_{E_k+F_k}) \circ L_k \\
&\subset \partial\Bigg(\sum_{k\in K} \iota_{E_k+F_k} \circ L_k\Bigg) \\
&= \partial\Bigg(\sum_{k\in K} \iota_{L_k^{-1}(E_k+F_k)}\Bigg) \\
&= \partial\iota_{\bigcap_{k\in K} L_k^{-1}(E_k+F_k)} \\
&= N_G.
\end{aligned} \tag{6.2}
$$

Thus, every solution to Problem 1.6 solves the variational inequality problem

$$
\text{find } \ x \in G \ \text{ such that } \quad (\forall y \in G)\Bigg(\exists\, x^* \in Ax + \sum_{i\in I} \partial(\phi_i \circ f_i)(x)\Bigg) \ \ \langle y - x, x^* + Cx + Qx\rangle \geqslant 0. \tag{6.3}
$$

The type of constraint set (6.1) appears in [16, 56]. To solve (6.3) via (5.11), we require the resolvents $J_{\mu_{k,n}B_k} = \text{proj}_{E_k}$, and $J_{\nu_{k,n}D_k} = \text{proj}_{F_k}$.

Example 6.2 Consider the implementation of Problem 1.6 in which

$$
A = \partial h_1, \ \ C = \nabla h_2, \ \ Q = 0, \ \text{ and } \ (\forall k \in K) \ \ B_k = \partial g_k \ \text{ and } \ D_k = \partial \ell_k, \tag{6.4}
$$

where $h_1 \in \Gamma_0(\mathcal{X})$, $h_2 \colon \mathcal{X} \to \mathbb{R}$ is a differentiable convex function with a Lipschitzian gradient, and for every $k \in K$, g_k and ℓ_k are functions in $\Gamma_0(\mathcal{Y}_k)$ such that $0 \in \text{sri}\,(\text{dom}\, g_k^* - \text{dom}\, \ell_k^*)$. Then, if $\overline{x} \in \mathcal{X}$ solves Problem 1.6, it solves

$$
\underset{x\in\mathcal{X}}{\text{minimize}} \ \ \sum_{i\in I} (\phi_i \circ f_i)(x) + \sum_{k\in K} (g_k \,\Box\, \ell_k)(L_k x) + h_1(x) + h_2(x). \tag{6.5}
$$

To solve this problem via (5.11), we require the resolvents $J_{\gamma_n A} = \text{prox}_{\gamma_n h_1}$, $J_{\mu_{k,n}B_k} = \text{prox}_{\mu_{k,n} g_k}$, and $J_{\nu_{k,n}D_k} = \text{prox}_{\nu_{k,n}\ell_k}$.

Proof. First, in view of [7, Corollaire 10], C is cocoercive. Next, for every $k \in K$, it follows from Lemma 2.7 that $\partial g_k \,\Box\, \partial \ell_k = \partial(g_k \,\Box\, \ell_k)$. Since $\overline{x} \in \mathcal{X}$ solves Problem 1.6, [58, Theorem 2.4.2(vii)–(viii)] yields

$$
\begin{aligned}
0 &\in \sum_{i\in I} \partial(\phi_i \circ f_i)(\overline{x}) + \sum_{k\in K} \big(L_k^* \circ (\partial g_k \,\Box\, \partial \ell_k) \circ L_k\big)\overline{x} + \partial h_1(\overline{x}) + \nabla h_2(\overline{x}) \\
&\subset \partial\Bigg(\sum_{i\in I} (\phi_i \circ f_i) + \sum_{k\in K} (g_k \,\Box\, \ell_k) \circ L_k + h_1 + h_2\Bigg)(\overline{x}),
\end{aligned} \tag{6.6}
$$

and the result follows from Lemma 2.5(iv). ☐

Remark 6.3 In Example 6.2, if I and K are singletons and $h_2 = 0$, we recover Problem 1.1, which can therefore be solved as an instance of (5.11) in Theorem 5.5.

The next example was suggested to us by Roberto Cominetti.

Example 6.4 Let X be a reflexive real Banach space, let J and K be disjoint finite sets, and let $\mathsf{h}_0 \in \Gamma_0(\mathsf{X})$. For every $j \in J$, let $0 < m_j \in \mathbb{N}$, let $\Phi_j \in \Gamma_0(\mathbb{R}^{m_j})$ be increasing in each component, let $(\mathsf{f}_{j,s})_{1 \leqslant s \leqslant m_j}$ be functions in $\Gamma_0(\mathsf{X})$, and let $\mathsf{F}_j : \mathsf{x} \mapsto (\mathsf{f}_{j,1}(\mathsf{x}), \ldots, \mathsf{f}_{j,m_j}(\mathsf{x}))$. Further, for every $k \in K$, let $\mathcal{Y}_k$ be a reflexive real Banach space, let g_k and ℓ_k be functions in $\Gamma_0(\mathcal{Y}_k)$ such that $0 \in \operatorname{sri}(\operatorname{dom} g_k^* - \operatorname{dom} \ell_k^*)$, and let $\mathsf{L}_k : \mathsf{X} \to \mathcal{Y}_k$ be linear and bounded. Consider the problem

$$\underset{\mathsf{x} \in \mathsf{X}}{\text{minimize}} \;\; \sum_{j \in J} (\Phi_j \circ \mathsf{F}_j)(\mathsf{x}) + \sum_{k \in K} (g_k \,\square\, \ell_k)(\mathsf{L}_k \mathsf{x}) + \mathsf{h}_0(\mathsf{x}). \tag{6.7}$$

To show that (6.7) is a special case of (6.5) in Example 6.2, let us first observe that, since $(\Phi_j)_{j \in J}$ are increasing componentwise,

$$\begin{aligned}(\forall j \in J)\big(\forall \mathsf{x} \in \mathsf{F}_j^{-1}(\operatorname{dom} \Phi_j)\big) \quad (\Phi_j \circ \mathsf{F}_j)(\mathsf{x}) &= \min_{\substack{(\zeta_s)_{1 \leqslant s \leqslant m_j} \in \mathbb{R}^{m_j} \\ (\forall s \in \{1,\ldots,m_j\})\ \mathsf{f}_{j,s}(\mathsf{x}) \leqslant \zeta_s}} \Phi_j(\zeta_1, \ldots, \zeta_{m_j}) \\ &= \min_{\mathsf{z} \in \mathbb{R}^{m_j}} \big(\iota_{]-\infty,0]^{m_j}}(\mathsf{F}_j(\mathsf{x}) - \mathsf{z}) + \Phi_j(\mathsf{z})\big).\end{aligned} \tag{6.8}$$

Therefore, (6.7) is equivalent to

$$\underset{\substack{\mathsf{x} \in \mathsf{X} \\ (\forall j \in J)\ \mathsf{z}_j \in \mathbb{R}^{m_j}}}{\text{minimize}} \;\; \sum_{j \in J} \iota_{]-\infty,0]^{m_j}}(\mathsf{F}_j(\mathsf{x}) - \mathsf{z}_j) + \sum_{j \in J} \Phi_j(\mathsf{z}_j) + \sum_{k \in K} (g_k \,\square\, \ell_k)(\mathsf{L}_k \mathsf{x}) + \mathsf{h}_0(\mathsf{x}). \tag{6.9}$$

Next, let us set $\mathcal{X} = \mathsf{X} \times \bigtimes_{j \in J} \mathbb{R}^{m_j}$ and $I = \big\{(j,s) \mid j \in J,\ 1 \leqslant s \leqslant m_j\big\}$. In addition, for every $j \in J$, let $\mathsf{z}_j = (\zeta_{j,s})_{1 \leqslant s \leqslant m_j} \in \mathbb{R}^{m_j}$ and, for every $i = (j,s) \in I$, let $\phi_i = \iota_{]-\infty,0]}$ and $f_i : (\mathsf{x}, (\mathsf{z}_j)_{j \in J}) \mapsto \mathsf{f}_{j,s}(\mathsf{x}) - \zeta_{j,s}$. Finally, set $h_1 : (\mathsf{x}, (\mathsf{z}_j)_{j \in J}) \mapsto \mathsf{h}_0(\mathsf{x}) + \sum_{j \in J} \Phi_j(\mathsf{z}_j)$, $h_2 = 0$, and, for every $k \in K$, $L_k : (\mathsf{x}, (\mathsf{z}_j)_{j \in J}) \mapsto \mathsf{L}_k \mathsf{x}$. Since the functions $(f_i)_{i \in I}$ and h_1 are in $\Gamma_0(\mathcal{X})$, (6.9) is a particular instance of (6.5). In connection with solving (6.7) via Theorem 5.5 when X is a Hilbert space, let us note that, for every $\mu \in \left]0,+\infty\right[$ and every $i = (j,s) \in I$, the computation of w_i is similar to that in Example 4.3, while

$$\operatorname{prox}_{\mu f_i} : \big(\mathsf{x}, (\mathsf{z}_{j'})_{j' \in J}\big) \mapsto \big(\operatorname{prox}_{\mu \mathsf{f}_{j,s}} \mathsf{x}, (\mathsf{z}_{j'} + \mu \delta_{j,j'} \mathsf{e}_{j,s})_{j' \in J}\big), \tag{6.10}$$

where $\mathsf{e}_{j,s}$ is the sth canonical vector of $\mathbb{R}^{m_j}$ and δ is the Kronecker symbol. Furthermore, for every $\mu \in \left]0,+\infty\right[$,

$$\operatorname{prox}_{\mu h_1} : \big(\mathsf{x}, (\mathsf{z}_j)_{j \in J}\big) \mapsto \big(\operatorname{prox}_{\mu \mathsf{h}_0} \mathsf{x}, (\operatorname{prox}_{\mu \Phi_j} \mathsf{z}_j)_{j \in J}\big). \tag{6.11}$$

Example 6.5 In Example 6.4, suppose that

$$(\forall j \in J) \quad \Phi_j : (\zeta_{j,s})_{1 \leqslant s \leqslant m_j} \mapsto \max_{1 \leqslant s \leqslant m_j} \zeta_{j,s}. \tag{6.12}$$

Then (6.7) reduces to

$$\underset{\mathsf{x} \in \mathsf{X}}{\text{minimize}} \;\; \sum_{j \in J} \max_{1 \leqslant s \leqslant m_j} \mathsf{f}_{j,s}(\mathsf{x}) + \sum_{k \in K} (g_k \,\square\, \ell_k)(\mathsf{L}_k \mathsf{x}) + \mathsf{h}_0(\mathsf{x}). \tag{6.13}$$

In a Hilbertian setting, using the numerical approach outlined in Theorem 5.5, we can solve (6.13) via an algorithm requiring only the proximity operators of the functions $\mathsf{f}_{j,s}$, g_k, ℓ_k, and h_0. This appears to be the first proximal algorithm to solve problems with such a mix of functions involving maxima. As seen in (6.11), to implement (5.11), we require the proximity operators of the functions in (6.12). These can be computed by using [8, Example 24.25].

References

[1] S. Adly, H. Attouch, and V. N. Vo, Newton-type inertial algorithms for solving monotone equations governed by sums of potential and nonpotential operators, *Appl. Math. Optim.*, vol. 85, art. 44, 2022.

[2] S. Alwadani, H. H. Bauschke, and X. Wang, Attouch-Théra duality, generalized cycles, and gap vectors, *SIAM J. Optim.*, vol. 31, pp. 1926–1946, 2021.

[3] H. Attouch, *Variational Convergence for Functions and Operators*. Pitman, Boston, MA, 1984.

[4] H. Attouch and H. Brezis, Duality for the sum of convex functions in general Banach spaces, in: *Aspects of Mathematics and Its Applications*, (A. Barroso, ed.), pp. 125–133. North-Holland, Amsterdam, 1986.

[5] H. Attouch, L. M. Briceño-Arias, and P. L. Combettes, A parallel splitting method for coupled monotone inclusions, *SIAM J. Control Optim.*, vol. 48, pp. 3246–3270, 2010.

[6] A. Auslender, R. Cominetti, and M. Haddou, Asymptotic analysis for penalty and barrier methods in convex and linear programming, *Math. Oper. Res.*, vol. 22, pp. 43–62, 1997.

[7] J.-B. Baillon and G. Haddad, Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones, *Israel J. Math.*, vol. 26, pp. 137–150, 1977.

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

[9] J. Bolte, Z. Chen, and E. Pauwels, The multiproximal linearization method for convex composite problems, *Math. Program.*, vol. A182, pp. 1–36, 2020.

[10] R. I. Boţ, *Conjugate Duality in Convex Optimization*. Springer, Berlin, 2010.

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

[12] L. M. Briceño-Arias and P. L. Combettes, A monotone+skew splitting model for composite monotone inclusions in duality, *SIAM J. Optim.*, vol. 21, pp. 1230–1250, 2011.

[13] L. M. Briceño-Arias and P. L. Combettes, Monotone operator methods for Nash equilibria in non-potential games, in: *Computational and Analytical Mathematics*, (D. Bailey et al., eds.), pp. 143–159. Springer, New York, 2013.

[14] L. M. Briceño-Arias and D. Davis, Forward-backward-half forward algorithm for solving monotone inclusions, *SIAM J. Optim.*, vol. 28, pp. 2839–2871, 2018.

[15] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.*, vol. 491, art. 124315, 2020.

[16] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Math. Oper. Res.*, vol. 47, pp. 1082–1109, 2022.

[17] J. V. Burke, Descent methods for composite nondifferentiable optimization problems, *Math. Program.*, vol. 33, pp. 260–279, 1985.

[18] J. V. Burke and A. Engle, Strong metric (sub)regularity of Karush-Kuhn-Tucker mappings for piecewise linear-quadratic convex-composite optimization and the quadratic convergence of Newton's method, *Math. Oper. Res.*, vol. 45, pp. 1164–1192, 2020.

[19] J. V. Burke, T. Hoheisel, and Q. V. Nguyen, A study of convex convex-composite functions via infimal convolution with applications, *Math. Oper. Res.*, vol. 46, pp. 1324–1348, 2021.

[20] V. Cevher and B. C. Vũ, A reflected forward-backward splitting method for monotone inclusions involving Lipschitzian operators, *Set-Valued Var. Anal.*, vol. 29, pp. 163–174, 2021.

[21] G. Chen and M. Teboulle, A proximal-based decomposition method for convex minimization problems, *Math. Program.*, vol. 64, pp. 81–101, 1994.

[22] C. Combari, M. Laghdir, and L. Thibault, Sous-différentiels de fonctions convexes composées, *Ann. Sci. Math. Québec*, vol. 18, pp. 119–148, 1994.

[23] C. Combari, M. Laghdir, and L. Thibault, A note on subdifferentials of convex composite functionals, *Arch. Math. (Basel)*, vol. 67, pp. 239–252, 1996.

[24] C. W. Combettes and S. Pokutta, Complexity of linear minimization and projection on some sets, *Oper. Res. Lett.*, vol. 49, pp. 565–571, 2021.

[25] P. L. Combettes and L. E. Glaudin, Proximal activation of smooth functions in splitting algorithms for convex image recovery, *SIAM J. Imaging Sci.*, vol. 12, pp. 1905–1935, 2019.

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

[27] P. L. Combettes and J.-C. Pesquet, Deep neural network structures solving variational inequalities, *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, 2020.

[28] P. L. Combettes and J.-C. Pesquet, Fixed point strategies in data science, *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

[29] L. Condat, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, *J. Optim. Theory Appl.*, vol. 158, pp. 460–479, 2013.

[30] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer, New York, 2003.

[31] M. Jaggi, Revisiting Frank-Wolfe: Projection-free sparse convex optimization, *Proc. 30th Int. Conf. Machine Learn.*, vol. 28, pp. 427–435, 2013.

[32] J. L. Joly and P. J. Laurent, Stability and duality in convex minimization problems, *Rev. Française Informat. Recherche Opérationnelle*, sér. R2, vol. 5, pp. 3–42, 1971.

[33] S. S. Kutateladze, Formulas for computing subdifferentials, *Soviet. Math. Dokl.*, vol. 18, pp. 146–148, 1977.

[34] B. Lemaire, Application of a subdifferenlial of a convex composite functional to optimal control in variational inequalities, *Lecture Notes in Economics and Mathematical Systems*, vol. 255, pp. 103–117, 1985.

[35] C. Lescarret, Sous-différentialité de fonctions composées, *Séminaire d'Analyse Unilatérale*, exposé no. 2, Montpellier, France, 1968.

[36] A. S. Lewis and S. J. Wright, A proximal method for composite minimization, *Math. Program.*, vol. 158, pp. 501–546, 2016.

[37] J.-L. Lions, *Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires.* Dunod, Paris, 1969.

[38] Yu. Malitsky and M. K. Tam, A forward-backward splitting method for monotone inclusions without cocoercivity, *SIAM J. Optim.*, vol. 30, pp. 1451–1472, 2020.

[39] B. S. Mordukhovich, N. M. Nam, and J. Salinas, Solving a generalized Heron problem by means of convex analysis, *Amer. Math. Monthly*, vol. 119, pp. 87–99, 2012.

[40] E. Pauwels, The value function approach to convergence analysis in composite optimization, *Oper. Res. Lett.*, vol. 44, pp. 790–795, 2016.

[41] T. Pennanen, Graph-convex mappings and K-convex functions, *J. Convex Anal.*, vol. 6, pp. 235–266, 1999.

[42] T. Pennanen, Dualization of generalized equations of maximal monotone type, *SIAM J. Optim.*, vol. 10, pp. 809–835, 2000.

[43] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes – The Art of Scientific Computing*, 3rd. ed. Cambridge University Press, Cambridge, MA, 2007.

[44] A. Repetti and Y. Wiaux, Variable metric forward-backward algorithm for composite minimization problems, *SIAM J. Optim.*, vol. 31, pp. 1215–1241, 2021.

[45] R. T. Rockafellar, Level sets and continuity of conjugate convex functions, *Trans. Amer. Math. Soc.*, vol. 123, pp. 46–63, 1966.

[46] R. T. Rockafellar, Duality and stability in extremum problems involving convex functions, *Pacific J. Math.*, vol. 21, pp. 167–187, 1967.

[47] R. T. Rockafellar, Convex functions and duality in optimization problems and dynamics, in: *Mathematical Systems Theory and Economics I*, (H. W. Kuhn and G. P. Szegö, eds.), pp. 117–141. Springer, New York, 1969.

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

[49] R. T. Rockafellar, Monotone operators associated with saddle-functions and minimax problems, in: *Nonlinear Functional Analysis, Part 1*, (F. E. Browder, ed.), pp. 241–250. AMS, Providence, RI, 1970.

[50] R. T. Rockafellar, *Conjugate Duality and Optimization*. SIAM, Philadelphia, PA, 1974.

[51] R. T. Rockafellar, Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality, *Math. Program.*, vol. A198, pp. 159–194, 2023.

[52] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, 3rd printing. Springer, New York, 2009.

[53] A. M. Rubinov, Sublinear operators and their applications, *Russian Math. Surveys*, vol. 32, pp. 115–175, 1977.

[54] E. K. Ryu and B. C. Vũ, Finding the forward-Douglas–Rachford-forward method, *J. Optim. Theory Appl.*, vol. 184, pp. 858–876, 2020.

[55] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.

[56] X. Wang, J. Zhang, and W. Zhang, The distance between convex sets with Minkowski sum structure: Application to collision detection, *Comput. Optim. Appl.*, vol. 77, pp. 465–490, 2020.

[57] E. Zeidler, *Nonlinear Functional Analysis and Its Applications II/B: Nonlinear Monotone Operators.* Springer, New York, 1990.

[58] C. Zălinescu, *Convex Analysis in General Vector Spaces.* World Scientific Publishing, River Edge, NJ, 2002.