TAIWANESE JOURNAL OF MATHEMATICS
Vol. 17, No. 6, pp. 1983-2009, December 2013
DOI: 10.11650/tjm.17.2013.3087
This paper is available online at http://journal.taiwanmathsoc.org.tw

SOLVING SYSTEMS OF MONOTONE INCLUSIONS VIA PRIMAL-DUAL SPLITTING TECHNIQUES

Radu Ioan Boţ[1], Ernö Robert Csetnek[2] and Erika Nagy[3]

Abstract. In this paper we propose an algorithm for solving systems of coupled monotone inclusions in Hilbert spaces. The operators arising in each of the inclusions of the system are processed in each iteration separately, namely, the single-valued are evaluated explicitly (forward steps), while the set-valued ones via their resolvents (backward steps). In addition, most of the steps in the iterative scheme can be executed simultaneously, this making the method applicable to a variety of convex minimization problems. The numerical performances of the proposed splitting algorithm are emphasized through applications in average consensus on colored networks and image classification via support vector machines.

1. Introduction and Problem Formulation

In recent years several splitting algorithms have emerged for solving monotone inclusion problems involving parallel sums and compositions with linear continuous operators, which eventually are reduced to finding the zeros of the sum of a maximally monotone operator and a cocoercive or a monotone and Lipschitz continuous operator. The later problems were solved by employing in an appropriate product space forward-backward or forward-backward-forward algorithms, respectively, and gave rise to so-called primal-dual splitting methods (see [11, 17, 14, 29] and the references therein).

Received March 25, 2013, accepted May 10, 2013.
Communicated by Jen-Chih Yao.
2010 *Mathematics Subject Classification*: 47H05, 65K05, 90C25, 90C46.
Key words and phrases: Convex minimization, Coupled systems, Forward-backward-forward algorithm, Monotone inclusion, Operator splitting.
[1]On leave to University of Vienna. Research partially supported by DFG (German Research Foundation), Project BO 2516/4-1.
[2]Research supported by DFG (German Research Foundation), project BO 2516/4-1.
[3]Research made during the stay of the author in the academic year 2012/2013 at Chemnitz University of Technology as a guest of the Chair of Applied Mathematics (Approximation Theory).
This work was supported by a grant of the Romanian National Authority for Scientific Research CNCS-UEFISCDI, project number PN-II-ID-PCE-2011-3-0024, by POSDRU/107/1.5/S/76841 - "Modern Doctoral Studies: internationalization and inter-disciplinarity", Romania and by Collegium Talentum, Hungary.

Recently, one can remark the interest of researchers in solving systems of monotone inclusion problems [1, 4, 16]. This is motivated by the fact that convex optimization problems arising, for instance, in areas like image processing [9], multifacility location problems [12, 21], average consensus in network coloring [22, 23] and support vector machines classification [19] are to be solved with respect to multiple variables, very often linked in different manners, for instance, by linear equations.

The present research is motivated by the investigations made in [1]. The authors propose there an algorithm for solving coupled monotone inclusion problems, where the variables are linked by some operators which satisfy jointly a cocoercivity property. The iterative scheme in [1] relies on a forward-backward algorithm applied in an appropriate product space and it is employed in the solving of a class of convex optimization problems with multiple variables where some of the functions involved need to be differentiable. Our aim is to overcome the necessity of having differentiability for some of the functions occurring in the objective of the convex optimization problems in [1]. To this end we consider first a more general system of monotone inclusions, for which the coupling operator satisfies a Lipschitz continuity property, along with its dual system of monotone inclusions in an extended sense of the Attouch-Théra duality (see [2]). The simultaneous solving of the primal and dual system of monotone inclusions is reduced to the problem of finding the zeros of the sum of a maximally monotone operator and a monotone and Lipschitz continuous operator in an appropriate product space. The latter problem is solved by a forward-backward-forward algorithm, fact that allows us to provide for the resulting iterative scheme, which proves to have a high parallelizable formulation, both weak and strong convergence assertions.

The problem under consideration is as follows.

Problem 1.1. Let $m \geq 1$ be a positive integer, $(\mathcal{H}_i)_{1\leq i\leq m}$ be real Hilbert spaces and for $i = 1, ..., m$ let $B_i : \mathcal{H}_1 \times \ldots \times \mathcal{H}_m \to \mathcal{H}_i$ be a μ_i-Lipschitz continuous operator with $\mu_i \in \mathbb{R}_{++}$ jointly satisfying the monotonicity property

$$(\forall (x_1, \ldots, x_m) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m)(\forall (y_1, \ldots, y_m) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m)$$

$$\sum_{i=1}^{m} \langle x_i - y_i | B_i(x_1, \ldots, x_m) - B_i(y_1, \ldots, y_m)\rangle_{\mathcal{H}_i} \geq 0. \tag{1.1}$$

For every $i = 1, \ldots, m$, let $\mathcal{G}_i$ be a real Hilbert space, $A_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ a maximally monotone operator, $C_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ a monotone operator such that C_i^{-1} is ν_i-Lipschitz continuous with $\nu_i \in \mathbb{R}_+$ and $L_i : \mathcal{H}_i \to \mathcal{G}_i$ a linear continuous operator. The problem is to solve the system of coupled inclusions (see (2.4) for the definition of the parallel sum of two operators)

$$\text{find } \overline{x}_1 \in \mathcal{H}_1, \ldots, \overline{x}_m \in \mathcal{H}_m \text{ such that}$$
$$\left\{\begin{array}{l} 0 \in L_1^*(A_1 \Box C_1)(L_1\overline{x}_1) + B_1(\overline{x}_1, \ldots, \overline{x}_m) \\ \quad\vdots \\ 0 \in L_m^*(A_m \Box C_m)(L_m\overline{x}_m) + B_m(\overline{x}_1, \ldots, \overline{x}_m) \end{array}\right. \tag{1.2}$$

together with its dual system

$$\begin{array}{c}\text{find } \overline{v}_1 \in \mathcal{G}_1, \ldots, \overline{v}_m \in \mathcal{G}_m \text{ such that} \\ (\exists x_1 \in \mathcal{H}_1, \ldots, \exists x_m \in \mathcal{H}_m)\end{array} \left\{\begin{array}{l} 0 = L_1^*\overline{v}_1 + B_1(x_1, \ldots, x_m) \\ \vdots \\ 0 = L_m^*\overline{v}_m + B_m(x_1, \ldots, x_m) \\ \overline{v}_1 \in (A_1 \Box C_1)(L_1x_1) \\ \vdots \\ \overline{v}_m \in (A_m \Box C_m)(L_mx_m) \end{array}\right. . \tag{1.3}$$

We say that $(\overline{x}_1, \ldots, \overline{x}_m, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m \times \mathcal{G}_1 \ldots \times \mathcal{G}_m$ is a primal-dual solution to Problem 1.1, if

$$0 = L_i^*\overline{v}_i + B_i(\overline{x}_1, \ldots, \overline{x}_m) \text{ and } \overline{v}_i \in (A_i \Box C_i)(L_i\overline{x}_i),\ i = 1, \ldots, m. \tag{1.4}$$

If $(\overline{x}_1, \ldots, \overline{x}_m, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m \times \mathcal{G}_1 \ldots \times \mathcal{G}_m$ is a primal-dual solution to Problem 1.1, then $(\overline{x}_1, \ldots, \overline{x}_m)$ is a solution to (1.2) and $(\overline{v}_1, \ldots, \overline{v}_m)$ is a solution to (1.3). Notice also that

$$(\overline{x}_1, \ldots, \overline{x}_m) \text{ solves (1.2)} \Leftrightarrow 0 \in L_i^*(A_i \Box C_i)(L_i\overline{x}_i) + B_i(\overline{x}_1, \ldots, \overline{x}_m),\ i = 1, \ldots, m \Leftrightarrow$$
$$\exists\, \overline{v}_1 \in \mathcal{G}_1, \ldots, \overline{v}_m \in \mathcal{G}_m \text{ such that } \left\{\begin{array}{l} 0 = L_i^*\overline{v}_i + B_i(\overline{x}_1, \ldots, \overline{x}_m),\ i = 1, \ldots, m \\ \overline{v}_i \in (A_i \Box C_i)(L_i\overline{x}_i),\ i = 1, \ldots, m. \end{array}\right. .$$

Thus, if $(\overline{x}_1, \ldots, \overline{x}_m)$ is a solution to (1.2), then there exists $(\overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{G}_1 \times \ldots \mathcal{G}_m$ such that $(\overline{x}_1, \ldots, \overline{x}_m, \overline{v}_1, \ldots, \overline{v}_m)$ is a primal-dual solution to Problem 1.1 and, if $(\overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{G}_1 \times \ldots \mathcal{G}_m$ is a solution to (1.3), then there exists $(\overline{x}_1, \ldots, \overline{x}_m) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m$ such that $(\overline{x}_1, \ldots, \overline{x}_m, \overline{v}_1, \ldots, \overline{v}_m)$ is a primal-dual solution to Problem 1.1.

The paper is organized as follows. In the next section we give some necessary notations and preliminary results in order to facilitate the reading of the manuscript. In Section 3 we formulate the primal-dual splitting algorithm for solving Problem 1.1 and investigate its convergence behaviour, while in Section 4 applications to solving primal-dual pairs of convex optimization problems are presented. Finally, in Section 5, we present two applications of the proposed algorithm addressing the average consensus problem on colored networks and the classification of images via support vector machines.

2. Notations and Preliminaries

Let us recall some elements of convex analysis and monotone operator theory which are needed in the sequel (see [3, 5, 6, 20, 30, 25]).

For every real Hilbert space occurring in the paper we generically denote its *inner product* with $\langle\cdot|\cdot\rangle$ and the associated *norm* with $\|\cdot\| = \sqrt{\langle\cdot|\cdot\rangle}$. In order to avoid confusion, when needed, appropriate indices for the inner product and norm are used. The symbols $\rightharpoonup$ and $\rightarrow$ denote weak and strong convergence, respectively. Further, $\mathbb{R}_+$ denotes the set of nonnegative real numbers, $\mathbb{R}_{++}$ the set of strictly positive real numbers and $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ the extended real-line.

Let $\mathcal{H}$ be a real Hilbert space. The *indicator function* $\delta_C : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ of a set $C \subseteq \mathcal{H}$ is defined by $\delta_C(x) = 0$ for $x \in C$ and $\delta_C(x) = +\infty$, otherwise. If C is convex, we denote by $\mathrm{sqri}C := \{x \in C : \cup_{\lambda>0}\lambda(C - x)$ is a closed linear subspace of $\mathcal{H}\}$ its *strong quasi-relative interior*. Notice that we always have $\mathrm{int}\, C \subseteq \mathrm{sqri}C$ (in general this inclusion may be strict). If $\mathcal{H}$ is finite-dimensional, then $\mathrm{sqri}C$ coincides with $\mathrm{ri}\, C$, the *relative interior* of C, which is the interior of C with respect to its affine hull.

For a function $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ we denote by $\mathrm{dom}\, f := \{x \in \mathcal{H} : f(x) < +\infty\}$ its *effective domain* and call f *proper* if $\mathrm{dom}\, f \neq \varnothing$ and $f(x) > -\infty$ for all $x \in \mathcal{H}$. We denote by $\Gamma(\mathcal{H})$ the set of proper, convex and lower semicontinuous functions $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$. The *conjugate function* of f is $f^* : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, $f^*(u) = \sup\{\langle u, x\rangle - f(x) : x \in \mathcal{H}\}$ for all $u \in \mathcal{H}$ and, if $f \in \Gamma(\mathcal{H})$, then $f^* \in \Gamma(\mathcal{H})$, as well. The function f is said to be γ-strongly convex with $\gamma > 0$, if $f - \gamma/2\|\cdot\|^2$ is convex. The *(convex) subdifferential* of the function f at $x \in \mathcal{H}$ is the set $\partial f(x) = \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\, \langle y - x \mid u\rangle + f(x) \leq f(y)\}$, if $f(x) \in \mathbb{R}$, and is taken to be the empty set, otherwise. The *infimal con volution* of two proper functions $f,\, g : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ is defined by

$$f \,\Box\, g : \mathcal{H} \rightarrow \overline{\mathbb{R}}, f \,\Box\, g(x) = \inf_{y\in\mathcal{H}} \{f(y) + g(x - y)\}.$$

When $f \in \Gamma(\mathcal{H})$ and $\gamma > 0$, for every $x \in \mathcal{H}$ we denote by $\mathrm{Prox}_{\gamma f}(x)$ the *proximal point* of parameter γ of f at x, which is the unique optimal solution of the optimization problem

$$\inf_{y\in\mathcal{H}} \left\{f(y) + \frac{1}{2\gamma}\|y - x\|^2\right\}. \tag{2.1}$$

We have *Moreau's decomposition formula*

$$\mathrm{Prox}_{\gamma f} + \gamma \mathrm{Prox}_{(1/\gamma)f^*} \circ \gamma^{-1}\mathrm{Id} = \mathrm{Id}, \tag{2.2}$$

where operator Id denotes the identity on the underlying Hilbert space.

Let $2^{\mathcal{H}}$ be the power set of $\mathcal{H}$, $M : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ a set-valued operator and $\gamma > 0$. We denote by $\mathrm{zer}\, M = \{x \in \mathcal{H} : 0 \in Mx\}$ the set of *zeros* of M and by gra $M = \{(x,u) \in \mathcal{H} \times \mathcal{H} : u \in Mx\}$ the *graph* of M. We say that the operator M is *monotone* if $\langle x-y|u-v\rangle \geq 0$ for all $(x,u), (y,v) \in \text{gra } M$ and it is said to be *maximally monotone* if there exists no monotone operator $N : \mathcal{H} \to 2^{\mathcal{H}}$ such that gra N properly contains gra M. The operator M is said to be *uniformly monotone* with modulus $\phi_M : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$, if ϕ_M is increasing, vanishes only at 0, and $\langle x-y|u-v\rangle \geq \phi_M(\|x-y\|)$ for all $(x,u), (y,v) \in \text{gra } M$. A prominent representative of the class of uniformly monotone operators are the strongly monotone operators. We say that M is γ-*strongly monotone*, if $\langle x-y, u-v\rangle \geq \gamma\|x-y\|^2$ for all $(x,u), (y,v) \in \text{gra } M$.

The *inverse* of M is $M^{-1} : \mathcal{H} \to 2^{\mathcal{H}}$, $u \mapsto \{x \in \mathcal{H} : u \in Mx\}$. The *resolvent* of an operator $M : \mathcal{H} \to 2^{\mathcal{H}}$ is $J_M : \mathcal{H} \to 2^{\mathcal{H}}$, $J_M = (\text{Id } + M)^{-1}$. If M is maximally monotone, then $J_M : \mathcal{H} \to \mathcal{H}$ is single-valued and maximally monotone (cf. [3, Proposition 23.7 and Corollary 23.10]). We have (see [3, Proposition 23.18])

$$J_{\gamma M} + \gamma J_{\gamma^{-1}M^{-1}} \circ \gamma^{-1}\text{Id } = \text{Id } . \tag{2.3}$$

A single-valued operator $M : \mathcal{H} \to \mathcal{H}$ is said to be γ-*cocoercive*, if $\langle x-y, Mx - My\rangle \geq \gamma\|Mx - My\|^2$ for all $(x,y) \in \mathcal{H}\times\mathcal{H}$, while M is γ-*Lipschitz continuous* (here we allow also $\gamma = 0$ in order to comprise also the zero operator), if $\|Mx - My\| \leq \gamma\|x-y\|$ for all $(x,y) \in \mathcal{H} \times \mathcal{H}$. Obviously, every γ-cocoercive operator is monotone and γ^{-1}-Lipschitz continuous.

The *parallel sum* of two set-valued operators $M, N : \mathcal{H} \to 2^{\mathcal{H}}$ is defined as

$$M \,\square\, N : \mathcal{H} \to 2^{\mathcal{H}}, M \,\square\, N = \left(M^{-1} + N^{-1}\right)^{-1}. \tag{2.4}$$

If $f \in \Gamma(\mathcal{H})$, then $\partial f : \mathcal{H} \to 2^{\mathcal{H}}$ is a maximally monotone operator and $(\partial f)^{-1} = \partial f^*$. Moreover, $J_{\gamma\partial f} = (\text{Id }_{\mathcal{H}} + \gamma\partial f)^{-1} = \text{Prox }_{\gamma f}$.

Finally, we notice that for $f = \delta_C$, where $C \subseteq \mathcal{H}$ is a nonempty convex and closed set, it holds

$$J_{\partial\delta_C} = \text{Prox }_{\delta_C} = P_C, \tag{2.5}$$

where $P_C : \mathcal{H} \to C$ denotes the *projection operator* on C (see [3, Example 23.3 and Example 23.4]).

When $\mathcal{G}$ is a another real Hilbert space and $L : \mathcal{H} \to \mathcal{G}$ is a linear continuous operator, then the *norm* of L is defined as $\|L\| = \sup\{\|Lx\| : x \in \mathcal{H}, \|x\| \leq 1\}$, while $L^* : \mathcal{G} \to \mathcal{H}$, defined by $\langle Lx|y\rangle = \langle x|L^*y\rangle$ for all $(x,y) \in \mathcal{H} \times \mathcal{G}$, denotes the *adjoint operator* of L.

3. The Primal-dual Splitting Algorithm

The aim of this section is to provide an algorithm for solving Problem 1.1 and to furnish weak and strong convergence results for the sequences generated by it. The proposed iterative scheme has the property that each single-valued operator is processed explicitly, while each set-valued operator is evaluated via its resolvent. Absolutely summable sequences make the algorithm error-tolerant.

Algorithm 3.1. For every $i = 1, \ldots, m$ let $(a_{1,i,n})_{n\geq 0}$, $(b_{1,i,n})_{n\geq 0}$, $(c_{1,i,n})_{n\geq 0}$ be absolutely summable sequences in $\mathcal{H}_i$ and $(a_{2,i,n})_{n\geq 0}$, $(b_{2,i,n})_{n\geq 0}$, $(c_{2,i,n})_{n\geq 0}$ absolutely summable sequences in $\mathcal{G}_i$. Furthermore, set

$$\beta = \max\left\{\sqrt{\sum_{i=1}^{m}\mu_i^2}, \nu_1, \ldots, \nu_m\right\} + \max_{i=1,\ldots,m}\|L_i\|, \tag{3.1}$$

let $\varepsilon \in]0, 1/(\beta+1)[$ and $(\gamma_n)_{n\geq 0}$ be a sequence in $[\varepsilon, (1-\varepsilon)/\beta]$. For every $i = 1, \ldots, m$ let the initial points $x_{i,0} \in \mathcal{H}_i$ and $v_{i,0} \in \mathcal{G}_i$ be chosen arbitrary and set

$$(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} \text{For } i = 1, \ldots, m \\ \quad y_{i,n} = x_{i,n} - \gamma_n(L_i^* v_{i,n} + B_i(x_{1,n}, \ldots, x_{m,n}) + a_{1,i,n}) \\ \quad w_{i,n} = v_{i,n} - \gamma_n(C_i^{-1} v_{i,n} - L_i x_{i,n} + a_{2,i,n}) \\ \quad p_{i,n} = y_{i,n} + b_{1,i,n} \\ \quad r_{i,n} = J_{\gamma_n A_i^{-1}} w_{i,n} + b_{2,i,n} \\ \quad q_{i,n} = p_{i,n} - \gamma_n(L_i^* r_{i,n} + B_i(p_{1,n}, \ldots, p_{m,n}) + c_{1,i,n}) \\ \quad s_{i,n} = r_{i,n} - \gamma_n(C_i^{-1} r_{i,n} - L_i p_{i,n} + c_{2,i,n}) \\ \quad x_{i,n+1} = x_{i,n} - y_{i,n} + q_{i,n} \\ \quad v_{i,n+1} = v_{i,n} - w_{i,n} + s_{i,n}. \end{array} \right.$$

The following theorem establishes the convergence of Algorithm 3.1 by showing that its iterative scheme can be reduced to the error-tolerant version of the forward-backward-forward algorithm of Tseng (see [28]) recently provided in [11].

Theorem 3.1. *Suppose that Problem 1.1 has a primal-dual solution. For the sequences generated by Algorithm 3.1 the following statements are true:*

(*i*) $(\forall i \in \{1, \ldots, m\})$ $\sum_{n\geq 0}\|x_{i,n} - p_{i,n}\|^2_{\mathcal{H}_i} < +\infty$ *and* $\sum_{n\geq 0}\|v_{i,n} - r_{i,n}\|^2_{\mathcal{G}_i} < +\infty$.

(*ii*) *There exists a primal-dual solution* $(\overline{x}_1, \ldots, \overline{x}_m, \overline{v}_1, \ldots, \overline{v}_m)$ *to Problem 1.1 such that:*

  (*a*) $(\forall i \in \{1, \ldots, m\})$ $x_{i,n} \rightharpoonup \overline{x}_i$, $p_{i,n} \rightharpoonup \overline{x}_i$, $v_{i,n} \rightharpoonup \overline{v}_i$ *and* $r_{i,n} \rightharpoonup \overline{v}_i$ *as* $n \to +\infty$.

  (*b*) *if* $C_i^{-1}, i = 1, ..., m$, *is uniformly monotone and there exists an increasing function* $\phi_B : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$ *vanishing only at 0 and fulfilling*

$$\begin{aligned} &(\forall (x_1, \ldots, x_m) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m)(\forall (y_1, \ldots, y_m) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m) \\ &\sum_{i=1}^{m}\langle x_i - y_i | B_i(x_1, \ldots, x_m) - B_i(y_1, \ldots, y_m)\rangle_{\mathcal{H}_i} \\ &\geq \phi_B\Big(\|(x_1, \ldots, x_m) - (y_1, \ldots, y_m)\|\Big), \end{aligned} \tag{3.2}$$

then $(\forall i \in \{1,\ldots,m\})$ $x_{i,n} \to \overline{x}_i$, $p_{i,n} \to \overline{x}_i$, $v_{i,n} \to \overline{v}_i$ *and* $r_{i,n} \to \overline{v}_i$ *as* $n \to +\infty$.

Proof. We introduce the real Hilbert space $\boldsymbol{\mathcal{H}} = \mathcal{H}_1 \times \ldots \times \mathcal{H}_m$ endowed with the inner product and associated norm defined for $\boldsymbol{x} = (x_1,\ldots,x_m)$, $\boldsymbol{y} = (y_1,\ldots,y_m) \in \boldsymbol{\mathcal{H}}$ as

$$\langle \boldsymbol{x}|\boldsymbol{y}\rangle_{\boldsymbol{\mathcal{H}}} = \sum_{i=1}^{m} \langle x_i|y_i\rangle_{\mathcal{H}_i} \text{ and } \|\boldsymbol{x}\|_{\boldsymbol{\mathcal{H}}} = \sqrt{\sum_{i=1}^{m} \|x_i\|^2_{\mathcal{H}_i}}, \tag{3.3}$$

respectively. Furthermore, we consider the real Hilbert space $\boldsymbol{\mathcal{G}} = \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$ endowed with inner product and associated norm defined for $\boldsymbol{v} = (v_1,\ldots,v_m)$, $\boldsymbol{w} = (w_1,\ldots,w_m) \in \boldsymbol{\mathcal{G}}$ as

$$\langle \boldsymbol{v}|\boldsymbol{w}\rangle_{\boldsymbol{\mathcal{G}}} = \sum_{i=1}^{m} \langle v_i|w_i\rangle_{\mathcal{G}_i} \text{ and } \|\boldsymbol{v}\|_{\boldsymbol{\mathcal{G}}} = \sqrt{\sum_{i=1}^{m} \|v_i\|^2_{\mathcal{G}_i}}, \tag{3.4}$$

respectively. Let us now consider the Hilbert space $\boldsymbol{\mathcal{K}} = \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{G}}$ endowed with the inner product and associated norm defined, for $(\boldsymbol{x},\boldsymbol{v}),(\boldsymbol{y},\boldsymbol{w}) \in \boldsymbol{\mathcal{K}}$, as

$$\langle (\boldsymbol{x},\boldsymbol{v})|(\boldsymbol{y},\boldsymbol{w})\rangle_{\boldsymbol{\mathcal{K}}} = \langle \boldsymbol{x}|\boldsymbol{y}\rangle_{\boldsymbol{\mathcal{H}}} + \langle \boldsymbol{v}|\boldsymbol{w}\rangle_{\boldsymbol{\mathcal{G}}} \text{ and } \|(\boldsymbol{x},\boldsymbol{v})\|_{\boldsymbol{\mathcal{K}}} = \sqrt{\|\boldsymbol{x}\|^2_{\boldsymbol{\mathcal{H}}} + \|\boldsymbol{v}\|^2_{\boldsymbol{\mathcal{G}}}}, \tag{3.5}$$

respectively. Consider the set-valued operator

$$\begin{aligned} \boldsymbol{A} : \boldsymbol{\mathcal{K}} &\to 2^{\boldsymbol{\mathcal{K}}}, \\ (x_1,\ldots,x_m,v_1,\ldots,v_m) &\mapsto (0,\ldots,0,A_1^{-1}v_1,\ldots,A_m^{-1}v_m) \end{aligned}$$

and the single-valued operator

$$\begin{aligned} \boldsymbol{B} : \boldsymbol{\mathcal{K}} &\to \boldsymbol{\mathcal{K}}, \\ (x_1,\ldots,x_m,v_1,\ldots,v_m) &\mapsto (L_1^* v_1 + B_1(x_1,\ldots,x_m),\ldots,L_m^* v_m + B_m(x_1,\ldots,x_m), \\ &\qquad C_1^{-1}v_1 - L_1 x_1,\ldots,C_m^{-1}v_m - L_m x_m) \end{aligned}$$

We set

$$\overline{\boldsymbol{x}} = (\overline{x}_1,\ldots,\overline{x}_m,\overline{v}_1,\ldots,\overline{v}_m). \tag{3.6}$$

Then

$$\overline{\boldsymbol{x}} \in \operatorname{zer}(\boldsymbol{A}+\boldsymbol{B}) \Leftrightarrow \quad \begin{cases} 0 = L_1^*\overline{v}_1 + B_1(\overline{x}_1,\ldots,\overline{x}_m) \\ \vdots \\ 0 = L_m^*\overline{v}_m + B_m(\overline{x}_1,\ldots,\overline{x}_m) \\ 0 \in A_1^{-1}\overline{v}_1 + C_1^{-1}\overline{v}_1 - L_1\overline{x}_1 \\ \vdots \\ 0 \in A_m^{-1}\overline{v}_m + C_m^{-1}\overline{v}_m - L_m\overline{x}_m \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} 0 = L_1^*\overline{v}_1 + B_1(\overline{x}_1,\ldots,\overline{x}_m) \\ \vdots \\ 0 = L_m^*\overline{v}_m + B_m(\overline{x}_1,\ldots,\overline{x}_m) \\ \overline{v}_1 \in (A_1^{-1} + C_1^{-1})^{-1}(L_1\overline{x}_1) \\ \vdots \\ \overline{v}_m \in (A_m^{-1} + C_m^{-1})^{-1}(L_m\overline{x}_m) \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} 0 = L_1^*\overline{v}_1 + B_1(\overline{x}_1,\ldots,\overline{x}_m) \\ \vdots \\ 0 = L_m^*\overline{v}_m + B_m(\overline{x}_1,\ldots,\overline{x}_m) \\ \overline{v}_1 \in (A_1 \,\square\, C_1)(L_1\overline{x}_1) \\ \vdots \\ \overline{v}_m \in (A_m \,\square\, C_m)(L_m\overline{x}_m) \end{cases}.$$

Consequently, $\overline{\boldsymbol{x}} = (\overline{x}_1,\ldots,\overline{x}_m,\overline{v}_1,\ldots,\overline{v}_m)$ is a zero of the sum $\boldsymbol{A}+\boldsymbol{B}$ if and only if $(\overline{x}_1,\ldots,\overline{x}_m,\overline{v}_1,\ldots,\overline{v}_m)$ is a primal-dual solution to Problem 1.1. As already noticed, in this case, $(\overline{x}_1,\ldots,\overline{x}_m)$ solves the primal system (1.2) and $(\overline{v}_1,\ldots,\overline{v}_m)$ solves its dual system (1.3). Therefore, in order to determine a primal-dual solution to Problem 1.1, it is enough to find a zero of $\boldsymbol{A}+\boldsymbol{B}$.

Further, we will determine the nature of the operators $\boldsymbol{A}$ and $\boldsymbol{B}$. Since the operators $A_i, i = 1,...,m$, are maximally monotone, $\boldsymbol{A}$ is maximally monotone, too (cf. [3, Proposition 20.22 and Proposition 20.23]). Furthermore, by [3, Proposition 23.16], for all $\gamma \in \mathbb{R}_{++}$ and all $(x_1,\ldots,x_m,v_1,\ldots,v_m) \in \boldsymbol{\mathcal{K}}$ we have

$$J_{\gamma\boldsymbol{A}}(x_1,\ldots,x_m,v_1,\ldots,v_m) = (x_1,\ldots,x_m,J_{\gamma A_1^{-1}}v_1,\ldots,J_{\gamma A_m^{-1}}v_m). \tag{3.7}$$

Coming now to $\boldsymbol{B}$, let us prove first that this operator is monotone. Let $(x_1,\ldots,x_m, v_1,\ldots,v_m)$ and $(y_1,\ldots,y_m,w_1,\ldots,w_m)$ be two points in $\boldsymbol{\mathcal{K}}$. Using (1.1) and the monotonicity of $C_i^{-1},\ i=1,\ldots,m$, we obtain

$$
\begin{aligned}
&\langle (x_1,\dots,x_m,v_1,\dots,v_m)-(y_1,\dots,y_m,w_1,\dots,w_m) |\\
&\boldsymbol{B}(x_1,\dots,x_m,v_1,\dots,v_m)-\boldsymbol{B}(y_1,\dots,y_m,w_1,\dots,w_m)\rangle_{\boldsymbol{\mathcal{K}}}\\
=\ &\langle (x_1-y_1,\dots,x_m-y_m,v_1-w_1,\dots,v_m-w_m) |\\
&\big(B_1(x_1,\dots,x_m)-B_1(y_1,\dots,y_m)+L_1^*(v_1-w_1),\dots,\\
&B_m(x_1,\dots,x_m)-B_m(y_1,\dots,y_m)+L_m^*(v_m-w_m),\\
&C_1^{-1}v_1-C_1^{-1}w_1-L_1(x_1-y_1),\dots,C_m^{-1}v_m-C_m^{-1}w_m-L_m(x_m-y_m)\big)\rangle_{\boldsymbol{\mathcal{K}}}\\
=\ &\sum_{i=1}^m \langle x_i-y_i | B_i(x_1,\dots,x_m)-B_i(y_1,\dots,y_m)\rangle_{\mathcal{H}_i}\\
&+\sum_{i=1}^m \left\langle v_i-w_i | C_i^{-1}v_i-C_i^{-1}w_i\right\rangle_{\mathcal{G}_i}\\
&+\sum_{i=1}^m \Big(\langle x_i-y_i | L_i^*(v_i-w_i)\rangle - \langle v_i-w_i | L_i(x_i-y_i)\rangle_{\mathcal{H}_i}\Big)\\
=\ &\sum_{i=1}^m \langle x_i-y_i | B_i(x_1,\dots,x_m)_{\mathcal{H}_i}-B_i(y_1,\dots,y_m)\rangle_{\mathcal{H}_i}\\
&+\sum_{i=1}^m \left\langle v_i-w_i | C_i^{-1}v_i-C_i^{-1}w_i\right\rangle_{\mathcal{G}_i}\\
\geq\ &0.
\end{aligned}
\tag{3.8}
$$

Further, we show that $\boldsymbol{B}$ is a Lipschitz continuous operator and consider to this end $(x_1,\dots,x_m,v_1,\dots,v_m),(y_1,\dots,y_m,w_1,\dots,w_m)\in\boldsymbol{\mathcal{K}}$. It holds

$$
\begin{aligned}
&\Big\|\boldsymbol{B}(x_1,\dots,x_m,v_1,\dots,v_m)-\boldsymbol{B}(y_1,\dots,y_m,w_1,\dots,w_m)\Big\|_{\boldsymbol{\mathcal{K}}}\\
=&\Big\|\Big((B_1(x_1,\dots,x_m)-B_1(y_1,\dots,y_m),\dots,B_m(x_1,\dots,x_m)-B_m(y_1,\dots,y_m),\\
&C_1^{-1}v_1-C_1^{-1}w_1,\dots,C_m^{-1}v_m-C_m^{-1}w_m\Big)\\
&+\Big(L_1^*(v_1-w_1),\dots,L_m^*(v_m-w_m),-L_1(x_1-y_1),\dots,-L_m(x_m-y_m)\Big)\Big\|_{\boldsymbol{\mathcal{K}}}\\
\leq&\Big\|\Big((B_1(x_1,\dots,x_m)-B_1(y_1,\dots,y_m),\dots,B_m(x_1,\dots,x_m)-B_m(y_1,\dots,y_m),\\
&C_1^{-1}v_1-C_1^{-1}w_1,\dots,C_m^{-1}v_m-C_m^{-1}w_m\Big)\Big\|_{\boldsymbol{\mathcal{K}}}\\
&+\Big\|L_1^*(v_1-w_1),\dots,L_m^*(v_m-w_m),-L_1(x_1-y_1),\dots,-L_m(x_m-y_m)\Big\|_{\boldsymbol{\mathcal{K}}}\\
=&\sqrt{\sum_{i=1}^m \|B_i(x_1,\dots,x_m)-B_i(y_1,\dots,y_m)\|_{\mathcal{H}_i}^2+\sum_{i=1}^m \left\|C_i^{-1}v_i-C_i^{-1}w_i\right\|_{\mathcal{G}_i}^2}
\end{aligned}
\tag{3.9}
$$

$$
\begin{aligned}
&+\sqrt{\sum_{i=1}^{m}\|L_i^*(v_i-w_i)\|^2_{\mathcal{H}_i}+\sum_{i=1}^{m}\|L_i(x_i-y_i)\|^2_{\mathcal{G}_i}}\\
\leq&\sqrt{\sum_{i=1}^{m}\Big(\mu_i^2\sum_{j=1}^{m}\|x_j-y_j\|^2_{\mathcal{H}_j}\Big)+\sum_{i=1}^{m}\nu_i^2\,\|v_i-w_i\|^2_{\mathcal{G}_i}}\\
&+\sqrt{\sum_{i=1}^{m}\|L_i\|^2\,\|v_i-w_i\|^2_{\mathcal{G}_i}+\sum_{i=1}^{m}\|L_i\|^2\,\|x_i-y_i\|^2_{\mathcal{H}_i}}\\
\leq&\sqrt{\Big(\sum_{i=1}^{m}\mu_i^2\Big)\Big(\sum_{i=1}^{m}\|x_i-y_i\|^2_{\mathcal{H}_i}\Big)+\max_{i=1,\ldots,m}\nu_i^2\sum_{i=1}^{m}\|v_i-w_i\|^2_{\mathcal{G}_i}}\\
&+\sqrt{\max_{i=1,\ldots,m}\|L_i\|^2\Big(\sum_{i=1}^{m}\|v_i-w_i\|^2_{\mathcal{G}_i}+\sum_{i=1}^{m}\|x_i-y_i\|^2_{\mathcal{H}_i}\Big)}\\
\leq&\beta\,\|(x_1,\ldots,x_m,v_1,\ldots,v_m)-(y_1,\ldots,y_m,w_1,\ldots,w_m)\|_{\boldsymbol{\mathcal{K}}},
\end{aligned}
$$

hence, $\boldsymbol{B}$ is β-Lipschitz continuous, where β is the constant defined in (3.1).

Setting

$$
(\forall n\geq 0)\begin{cases}\boldsymbol{x}_n=(x_{1,n},\ldots,x_{m,n},v_{1,n},\ldots,v_{m,n})\\ \boldsymbol{y}_n=(y_{1,n},\ldots,y_{m,n},w_{1,n},\ldots,w_{m,n})\\ \boldsymbol{p}_n=(p_{1,n},\ldots,p_{m,n},r_{1,n},\ldots,r_{m,n})\\ \boldsymbol{q}_n=(q_{1,n},\ldots,q_{m,n},s_{1,n},\ldots,s_{m,n})\end{cases}
$$

and

$$
(\forall n\geq 0)\begin{cases}\boldsymbol{a}_n=(a_{1,1,n},\ldots,a_{1,m,n},a_{2,1,n},\ldots,a_{2,m,n})\\ \boldsymbol{b}_n=(b_{1,1,n},\ldots,b_{1,m,n},b_{2,1,n},\ldots,b_{2,m,n})\\ \boldsymbol{c}_n=(c_{1,1,n},\ldots,c_{1,m,n},c_{2,1,n},\ldots,c_{2,m,n})\end{cases},
$$

the summability hypotheses imply that

$$
\sum_{n\geq 0}\|\boldsymbol{a}_n\|_{\boldsymbol{\mathcal{K}}}<+\infty,\quad \sum_{n\geq 0}\|\boldsymbol{b}_n\|_{\boldsymbol{\mathcal{K}}}<+\infty\quad\text{and}\quad \sum_{n\geq 0}\|\boldsymbol{c}_n\|_{\boldsymbol{\mathcal{K}}}<+\infty. \tag{3.10}
$$

Furthermore, it follows that the iterative scheme in Algorithm 3.1 can be written as

$$
(\forall n\geq 0)\quad\begin{cases}\boldsymbol{y}_n=\boldsymbol{x}_n-\gamma_n(\boldsymbol{B}\boldsymbol{x}_n+\boldsymbol{a}_n)\\ \boldsymbol{p}_n=J_{\gamma_n\boldsymbol{A}}\boldsymbol{y}_n+\boldsymbol{b}_n\\ \boldsymbol{q}_n=\boldsymbol{p}_n-\gamma_n(\boldsymbol{B}\boldsymbol{p}_n+\boldsymbol{c}_n)\\ \boldsymbol{x}_{n+1}=\boldsymbol{x}_n-\boldsymbol{y}_n+\boldsymbol{q}_n.\end{cases} \tag{3.11}
$$

thus, it has the structure of the error-tolerant forward-backward-forward algorithm given in [11].

(i) It follows from [11, Theorem 2.5(i)] that

$$\sum_{n\geq 0} \|\boldsymbol{x}_n - \boldsymbol{p}_n\|^2_{\boldsymbol{\mathcal{K}}} < +\infty. \tag{3.12}$$

This means that

$$\begin{aligned}\sum_{n\geq 0} \|\boldsymbol{x}_n - \boldsymbol{p}_n\|^2_{\boldsymbol{\mathcal{K}}} &= \sum_{n\geq 0}\sum_{i=1}^{m}\Big(\|x_{i,n} - p_{i,n}\|^2_{\mathcal{H}_i} + \|v_{i,n} - r_{i,n}\|^2_{\mathcal{G}_i}\Big) \\ &= \sum_{i=1}^{m}\sum_{n\geq 0}\|x_{i,n} - p_{i,n}\|^2_{\mathcal{H}_i} + \sum_{i=1}^{m}\sum_{n\geq 0}\|v_{i,n} - r_{i,n}\|^2_{\mathcal{G}_i} < +\infty. \end{aligned}\tag{3.13}$$

Hence

$$\begin{aligned}&(\forall i \in \{1,\dots,m\})\ \sum_{n\geq 0}\|x_{i,n} - p_{i,n}\|^2_{\mathcal{H}_i} < +\infty \quad \text{and} \\ &\sum_{n\geq 0}\|v_{i,n} - r_{i,n}\|^2_{\mathcal{G}_i} < +\infty.\end{aligned}\tag{3.14}$$

(ii) It follows from [11, Theorem 2.5(ii)] that there exists an element $\overline{\boldsymbol{x}} = (\overline{x}_1, \dots, \overline{x}_m, \overline{v}_1, \dots, \overline{v}_m)$ in the set zer $(\boldsymbol{A} + \boldsymbol{B})$, thus a primal-dual solution to Problem 1.1, such that

$$\boldsymbol{x}_n \rightharpoonup \overline{\boldsymbol{x}} \quad \text{and} \quad \boldsymbol{p}_n \rightharpoonup \overline{\boldsymbol{x}}. \tag{3.15}$$

(ii)(a) It is a direct consequence of (3.15).

(ii)(b) Let be $i \in \{1, ..., m\}$. Since C_i^{-1} is uniformly monotone, there exists an increasing function $\phi_{C_i^{-1}} : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$, vanishing only at 0, such that

$$\langle x - y, v - w\rangle_{\mathcal{G}_i} \geq \phi_{C_i^{-1}}(\|x - y\|_{\mathcal{G}_i}) \ \ \forall (x, v), (y, w) \in \text{gra } C_i^{-1}. \tag{3.16}$$

Taking into consideration (3.2), we define the function $\phi_{\boldsymbol{B}} : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$,

$$\phi_{\boldsymbol{B}}(c) = \inf\Big\{\phi_B(a) + \sum_{i=1}^{m}\phi_{C_i^{-1}}(b_i) : \sqrt{a^2 + \sum_{i=1}^{m} b_i^2} = c\Big\}, \tag{3.17}$$

which is increasing, it vanishes only at 0 and it fulfills due to (3.8) the following inequality

$$
\begin{aligned}
&\langle (x_1,\ldots,x_m,v_1,\ldots,v_m)-(y_1,\ldots,y_m,w_1,\ldots,w_m)|\\
&\boldsymbol{B}(x_1,\ldots,x_m,v_1,\ldots,v_m)-\boldsymbol{B}(y_1,\ldots,y_m,w_1,\ldots,w_m)\rangle_{\boldsymbol{\mathcal{K}}}\\
=&\sum_{i=1}^m \langle x_i-y_i|B_i(x_1,\ldots,x_m)-B_i(y_1,\ldots,y_m)\rangle_{\mathcal{H}_i}\\
&+\sum_{i=1}^m \left\langle v_i-w_i|C_i^{-1}v_i-C_i^{-1}w_i\right\rangle_{\mathcal{G}_i}\\
\geq&\phi_B\Big(\,\|(x_1,\ldots,x_m)-(y_1,\ldots,y_m)\|_{\boldsymbol{\mathcal{H}}}\Big)+\sum_{i=1}^m \phi_{C_i^{-1}}\Big(\,\|v_i-w_i\|_{\mathcal{G}_i}\Big)\\
\geq&\phi_{\boldsymbol{B}}\Big(\,\|\boldsymbol{x}-\boldsymbol{y}\|_{\boldsymbol{\mathcal{K}}}\Big),\forall \boldsymbol{x}=(x_1,\ldots,x_m,v_1,\ldots,v_m),\boldsymbol{y}\\
=&(y_1,\ldots,y_m,w_1,\ldots,w_m)\in\boldsymbol{\mathcal{K}}.
\end{aligned}
\tag{3.18}
$$

Consequently, $\boldsymbol{B}$ is uniformly monotone and, according to [11, Theorem 2.5(iii)(b)], it follows that $\boldsymbol{x}_n \to \overline{\boldsymbol{x}}$ and $\boldsymbol{p}_n \to \overline{\boldsymbol{x}}$ as $n \to +\infty$. This leads to the desired conclusion. ■

4. Applications to Convex Minimization Problems

In this section we turn our attention to the solving of convex minimization problems with multiple variables via the primal-dual algorithm presented and investigated in this paper.

Problem 4.1. Let $m \geq 1$ and $p \geq 1$ be positive integers, $(\mathcal{H}_i)_{1\leq i\leq m}$, $(\mathcal{H}_i^{'})_{1\leq i\leq m}$ and $(\mathcal{G}_j)_{1\leq j\leq p}$ be real Hilbert spaces, $f_i, h_i \in \Gamma(\mathcal{H}_i^{'})$ such that h_i is ν_i^{-1}-strongly convex with $\nu_i \in \mathbb{R}_{++}, i=1,\ldots,m$, and $g_j \in \Gamma(\mathcal{G}_j)$ for $i=1,\ldots,m, j=1,\ldots,p$. Further, let be $K_i : \mathcal{H}_i \to \mathcal{H}_i^{'}$ and $L_{ji} : \mathcal{H}_i \to \mathcal{G}_j, i=1,\ldots,m, j=1,\ldots,p$ linear continuous operators. Consider the convex optimization problem

$$
\inf_{(x_1,\ldots,x_m)\in\mathcal{H}_1\times\ldots\times\mathcal{H}_m}\left\{\sum_{i=1}^m (f_i\,\Box\, h_i)(K_ix_i)+\sum_{j=1}^p g_j\left(\sum_{i=1}^m L_{ji}x_i\right)\right\}. \tag{4.1}
$$

In what follows we show that under an appropriate qualification condition solving the convex optimization problem (4.1) can be reduced to the solving of a system of monotone inclusions of type (1.2).

Define the following proper convex and lower semicontinuous function (see [3, Corollary 11.16 and Proposition 12.14])

$$
f:\mathcal{H}_1^{'}\times\ldots\times\mathcal{H}_m^{'}\to\overline{\mathbb{R}},\ \ (y_1,\ldots,y_m)\mapsto\sum_{i=1}^m (f_i\,\Box\, h_i)(y_i),
$$

and the linear continuous operator

$$K : \mathcal{H}_1 \times \ldots \times \mathcal{H}_m \to \mathcal{H}_1' \times \ldots \times \mathcal{H}_m', \quad (x_1, \ldots, x_m) \mapsto (K_1x_1, \ldots, K_mx_m),$$

having as adjoint

$$K^* : \mathcal{H}_1' \times \ldots \times \mathcal{H}_m' \to \mathcal{H}_1 \times \ldots \times \mathcal{H}_m, \quad (y_1, \ldots, y_m) \mapsto (K_1^*y_1, \ldots, K_m^*y_m).$$

Further, consider the linear continuous operators

$$L_j : \mathcal{H}_1 \times \ldots \times \mathcal{H}_m \to \mathcal{G}_j, \quad (x_1, \ldots, x_m) \mapsto \sum_{i=1}^{m} L_{ji}x_i, j = 1, ..., p,$$

having as adjoints

$$L_j^* : \mathcal{G}_j \to \mathcal{H}_1 \times \ldots \times \mathcal{H}_m, \quad y \mapsto (L_{j1}^*y, \ldots, L_{jm}^*y), j = 1, ..., p,$$

respectively. We have

$$\begin{aligned} &(\overline{x}_1, \ldots, \overline{x}_m) \text{ is an optimal solution to (4.1)} \\ &\Leftrightarrow (0, \ldots, 0) \in \partial \left(f \circ K + \sum_{j=1}^{p} g_j \circ L_j \right) (\overline{x}_1, \ldots \overline{x}_m). \end{aligned} \tag{4.2}$$

In order to split the above subdifferential in a sum of subdifferentials a so-called *qualification condition* must be fulfilled. In this context, we consider the following *interiority-type* qualification conditions:

$$(\mathrm{QC}_1) \left| \begin{array}{l} \text{there exists } x_i' \in \mathcal{H}_i \text{ such that} \\ K_ix_i' \in (\mathrm{dom}\ f_i + \mathrm{dom}\ h_i) \text{ and } f_i \,\Box\, h_i \text{ is continuous at } K_ix_i', i = 1, ..., m, \\ \text{and } \sum_{i=1}^m L_{ji}x_i' \in \mathrm{dom}\ g_j \text{ and } g_j \text{ is continuous at } \sum_{i=1}^m L_{ji}x_i', j = 1, ..., p \end{array} \right.$$

and

$$(\mathrm{QC}_2) \left| \begin{array}{l} (0, \ldots, 0) \in \ \mathrm{sqri} \Big(\prod_{i=1}^m (\mathrm{dom}\ f_i + \mathrm{dom}\ h_i) \times \prod_{j=1}^p \mathrm{dom}\ g_j \\ -\{(K_1x_1, \ldots, K_mx_m, \sum_{i=1}^m L_{1i}x_i, \ldots, \sum_{i=1}^m L_{pi}x_i) : \\ (x_1, \ldots, x_m) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m\} \Big). \end{array} \right.$$

We notice that $(QC_1) \Rightarrow (QC_2)$, these implications being in general strict, and refer the reader to [3, 5, 6, 20, 27, 30] and the references therein for other qualification conditions in convex optimization.

Remark 4.1. As already pointed out, due to [3, Corollary 11.16 and Proposition 12.14], for $i = 1, ..., m$, $f_i \,\Box\, h_i \in \Gamma(\mathcal{H}_i')$, hence, it is continuous on $\mathrm{int}\,(\mathrm{dom}\ f_i + \mathrm{dom}\ h_i)$, providing this set is nonempty (see [20, 30]). For other results regarding the continuity of the infimal convolution of convex functions we invite the reader to consult [26].

Remark 4.2. In finite-dimensional spaces the qualification condition (QC_2) is equivalent to

$$(\mathrm{QC}_2)\left|\begin{array}{l}\text{there exists } x_i' \in \mathcal{H}_i \text{ such that } K_i x_i' \in \operatorname{ri}\operatorname{dom} f_i + \operatorname{ri}\operatorname{dom} h_i, i = 1, ..., m,\\ \text{ and } \sum_{i=1}^m L_{ji} x_i' \in \operatorname{ri}\operatorname{dom} g_j, j = 1, ..., p.\end{array}\right.$$

Assuming that one of the qualification conditions above is fulfilled, we have that

$$\begin{aligned}
&(\overline{x}_1, \ldots, \overline{x}_m) \text{ is an optimal solution to (4.1)}\\
&\Leftrightarrow (0, \ldots, 0) \in K^* \partial f\big(K(\overline{x}_1, \ldots \overline{x}_m)\big) + \sum_{j=1}^p L_j^* \partial g_j\big(L_j(\overline{x}_1, \ldots \overline{x}_m)\big)\\
&\Leftrightarrow (0, \ldots, 0) \in \Big(K_1^* \partial(f_1 \,\Box\, h_1)(K_1 \overline{x}_1), \ldots, K_m^* \partial(f_m \,\Box\, h_m)(K_m \overline{x}_m)\Big)\\
&\qquad\qquad + \sum_{j=1}^p L_j^* \partial g_j\big(L_j(\overline{x}_1, \ldots \overline{x}_m)\big).
\end{aligned} \tag{4.3}$$

The strong convexity of the functions h_i imply that $\operatorname{dom} h_i^* = \mathcal{H}_i'$ (see [3, Corollary 11.16, Proposition 14.15]) and so $\partial(f_i \,\Box\, h_i) = \partial f_i \,\Box\, \partial h_i, i = 1, ..., m,$ (see [3, Proposition 24.27]). Thus, (4.3) is further equivalent to

$$(0, \ldots, 0) \in \Big(K_1^*(\partial f_1 \,\Box\, \partial h_1)(K_1 \overline{x}_1), \ldots, K_m^*(\partial f_m \,\Box\, \partial h_m)(K_m \overline{x}_m)\Big) + \sum_{j=1}^p L_j^* v_j,$$

where

$$\overline{v}_j \in \partial g_j\big(L_j(\overline{x}_1, \ldots \overline{x}_m)\big) \Leftrightarrow \overline{v}_j \in \partial g_j\Big(\sum_{i=1}^m L_{ji} \overline{x}_i\Big) \Leftrightarrow \sum_{i=1}^m L_{ji} \overline{x}_i \in \partial g_j^*(\overline{v}_j), j = 1, ..., p.$$

Then $(\overline{x}_1, \ldots, \overline{x}_m)$ is an optimal solution to (4.1) if and only if $(\overline{x}_1, \ldots, \overline{x}_m, \overline{v}_1, \ldots, \overline{v}_p)$ is a solution to

$$\left\{\begin{array}{l}
0 \in K_1^*(\partial f_1 \,\Box\, \partial h_1)(K_1 \overline{x}_1) + \sum_{j=1}^p L_{j1}^* \overline{v}_j\\
\quad\vdots\\
0 \in K_m^*(\partial f_m \,\Box\, \partial h_m)(K_m \overline{x}_m) + \sum_{j=1}^p L_{jm}^* \overline{v}_j\\
0 \in \partial g_1^*(\overline{v}_1) - \sum_{i=1}^m L_{1i} \overline{x}_i\\
\quad\vdots\\
0 \in \partial g_p^*(\overline{v}_p) - \sum_{i=1}^m L_{pi} \overline{x}_i.
\end{array}\right. \tag{4.4}$$

One can see now that (4.4) is a system of coupled inclusions of type (1.2), by taking

$$A_i = \partial f_i,\ C_i = \partial h_i,\ L_i = K_i, i = 1, ..., m,$$

$$A_{m+j} = \partial g_j^*,\ C_{m+j}(x) = \begin{cases} \mathcal{G}_j,\ x = 0 \\ \emptyset,\ \text{otherwise} \end{cases}, L_{m+j} = \mathrm{Id}\,_{\mathcal{G}_j}, j = 1, ..., p,$$

and, for $(x_1, ..., x_m, v_1, ..., v_p) \in \mathcal{H}_1 \times \ldots \mathcal{H}_m \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_p$, as coupling operators

$$B_i(x_1, \ldots, x_m, v_1, \ldots, v_p) = \sum_{j=1}^{p} L_{ji}^* v_j, i = 1, \ldots, m,$$

and

$$B_{m+j}(x_1, \ldots, x_m, v_1, \ldots, v_p) = -\sum_{i=1}^{m} L_{ji} x_i, j = 1, \ldots, p.$$

Define

$$B(x_1, \ldots, x_m, v_1, \ldots, v_p) = (B_1, \ldots, B_{m+p})(x_1, \ldots, x_m, v_1, \ldots, v_p) \tag{4.5}$$

$$= \left(\sum_{j=1}^{p} L_{j1}^* v_j, \ldots, \sum_{j=1}^{p} L_{jm}^* v_j, -\sum_{i=1}^{m} L_{1i} x_i, \ldots, -\sum_{i=1}^{m} L_{pi} x_i \right). \tag{4.6}$$

According to [3, Proposition 17.26(i), Corollary 16.24 and Theorem 18.15)] it follows that $C_i^{-1} = (\partial h_i)^{-1} = \partial h_i^* = \{\nabla h_i^*\}$ is ν_i-Lipschitz continuous for $i = 1, ..., m$. On the other hand, C_{m+j}^{-1} is the zero operator for $j = 1, ..., p$, thus 0-Lipschitz continuous.

Furthermore, the operators $B_i, i = 1, ..., m+p$ are linear and Lipschitz continuous, having as Lipschitz constants

$$\mu_i = \sqrt{\sum_{j=1}^{p} \|L_{ji}\|^2},\ i = 1, ..., m, \text{ and } \mu_{m+j} = \sqrt{\sum_{i=1}^{m} \|L_{ji}\|^2},\ j = 1, ..., p,$$

respectively. For every $(x_1, \ldots, x_m, v_1, \ldots, v_p), (y_1, \ldots, y_m, w_1, \ldots, w_p) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_p$ it holds

$$\sum_{i=1}^{m} \langle x_i - y_i | B_i(x_1, \ldots, x_m, v_1, \ldots, v_p) - B_i(y_1, \ldots, y_m, w_1, \ldots, w_p) \rangle_{\mathcal{H}_i}$$

$$+ \sum_{j=1}^{p} \langle v_j - w_j | B_{m+j}(x_1, \ldots, x_m, v_1, \ldots, v_p) - B_{m+j}(y_1, \ldots, y_m, w_1, \ldots, w_p) \rangle_{\mathcal{G}_j}$$

$$= \sum_{i=1}^{m} \left\langle x_i - y_i | \sum_{j=1}^{p} L_{ji}^* v_j - \sum_{j=1}^{p} L_{ji}^* w_j \right\rangle_{\mathcal{H}_i}$$

$$-\sum_{j=1}^{p}\left\langle v_j - w_j \middle| \sum_{i=1}^{m} L_{ji}x_i - \sum_{i=1}^{m} L_{ji}y_i \right\rangle_{\mathcal{G}_j} = 0,$$

thus (1.1) is fulfilled. This proves also that the linear continuous operator B is skew (i.e. $B^* = -B$).

Remark 4.3. Due to the fact that the operator B is skew, it is not cocoercive, hence, the approach presented in [1] cannot be applied in this context. On the other hand, in the light of the characterization given in (4.2), in order to determine an optimal solution of the optimization problem (4.1) (and an optimal solution of its Fenchel-type dual as well) one can use the primal-dual proximal splitting algorithms which have been recently introduced in [17, 29]. These approaches have the particularity to deal in an efficient way with sums of compositions of proper, convex and lower semicontinuous function with linear continuous operators, by evaluating separately each function via a backward step and each linear continuous operator (and its adjoint) via a forward step. However, the iterative scheme we propose in this section for solving (4.1) has the advantage of exploiting the separable structure of the problem.

Let us also mention that the dual inclusion problem of (4.4) reads (see (1.3))

$$\text{(4.7)} \quad \begin{array}{l} \text{find } \overline{w}_1 \in \mathcal{H}_1^{'}, \ldots, \overline{w}_m \in \mathcal{H}_m^{'}, \\ \overline{w}_{m+1} \in \mathcal{G}_1, \ldots, \overline{w}_{m+p} \in \mathcal{G}_p \text{ such that} \\ (\exists x_1 \in \mathcal{H}_1, \ldots, \exists x_m \in \mathcal{H}_m, \\ \exists v_1 \in \mathcal{G}_1, \ldots, \exists v_p \in \mathcal{G}_p) \end{array} \quad \left\{ \begin{array}{l} 0 = K_1^* \overline{w}_1 + \sum_{j=1}^{p} L_{j1}^* v_j \\ \vdots \\ 0 = K_m^* \overline{w}_m + \sum_{j=1}^{p} L_{jm}^* v_j \\ 0 = \overline{w}_{m+1} - \sum_{i=1}^{m} L_{1i} x_i \\ \vdots \\ 0 = \overline{w}_{m+p} - \sum_{i=1}^{m} L_{pi} x_i \\ \overline{w}_1 \in (\partial f_1 \Box \partial h_1)(K_1 x_1) \\ \vdots \\ \overline{w}_m \in (\partial f_m \Box \partial h_m)(K_m x_m) \\ \overline{w}_{m+1} \in \partial g_1^*(v_1) \\ \vdots \\ \overline{w}_{m+p} \in \partial g_p^*(v_p). \end{array} \right.$$

Then $(\overline{x}_1, ..., \overline{x}_m, \overline{v}_1, ..., \overline{v}_p, \overline{w}_1, ..., \overline{w}_m, \overline{w}_{m+1}, ..., \overline{w}_{m+p})$ is a primal-dual solution to (4.4) - (4.7), if

$$\overline{w}_i \in (\partial f_i \,\square\, \partial h_i)(K_i \overline{x}_i), \overline{w}_{m+j} \in \partial g_j^*(\overline{v}_j),$$
$$0 = K_i^* \overline{w}_i + \sum_{j=1}^{p} L_{ji}^* \overline{v}_j \text{ and } 0 = \overline{w}_{m+j} - \sum_{i=1}^{m} L_{ji} \overline{x}_i, i = 1, ..., m, j = 1, ..., p.$$

Provided that $(\overline{x}_1, ..., \overline{x}_m, \overline{v}_1, ..., \overline{v}_p, \overline{w}_1, ..., \overline{w}_m, \overline{w}_{m+1}, ..., \overline{w}_{m+p})$ is a primal-dual solution to (4.4) - (4.7), it follows that $(\overline{x}_1, ...\overline{x}_m)$ is an optimal solution to (4.1) and $(\overline{w}_1, ..., \overline{w}_m, \overline{v}_1, ..., \overline{v}_p)$ is an optimal solution to its *Fenchel-type* dual problem
(4.8)

$$\sup_{\substack{(w_1,...,w_m,w_{m+1},...,w_{m+p}) \in \mathcal{H}_1' \times ... \times \mathcal{H}_m' \times \mathcal{G}_1 \times ... \times \mathcal{G}_p \\ K_i^* w_i + \sum_{j=1}^{p} L_{ji}^* w_{m+j} = 0, i=1,...,m}} \left\{ -\sum_{i=1}^{m} \left(f_i^*(w_i) + h_i^*(w_i)\right) - \sum_{j=1}^{p} g_j^*(w_{m+j}) \right\}.$$

Algorithm 3.1 gives rise to the following iterative scheme for solving (4.4) - (4.7).

Algorithm 4.2.
For every $i = 1, \ldots, m$ and every $j = 1, \ldots, p$ let $(a_{1,i,n})_{n \geq 0}$, $(b_{1,i,n})_{n \geq 0}$, $(c_{1,i,n})_{n \geq 0}$, be absolutely summable sequences in $\mathcal{H}_i$, $(a_{2,i,n})_{n \geq 0}$, $(b_{2,i,n})_{n \geq 0}$, $(c_{2,i,n})_{n \geq 0}$ be absolutely summable sequences in $\mathcal{H}_i'$ and $(a_{1,m+j,n})_{n \geq 0}$, $(a_{2,m+j,n})_{n \geq 0}$, $(b_{1,m+j,n})_{n \geq 0}$, $(b_{2,m+j,n})_{n \geq 0}$, $(c_{1,m+j,n})_{n \geq 0}$ and $(c_{2,m+j,n})_{n \geq 0}$ be absolutely summable sequences in $\mathcal{G}_j$. Furthermore, set

$$\text{(4.9)} \qquad \beta = \max \left\{ \sqrt{\sum_{i=1}^{m+p} \mu_i^2}, \nu_1, \ldots, \nu_m \right\} + \max \left\{ \|K_1\|, \ldots, \|K_m\|, 1 \right\},$$

where

$$\text{(4.10)} \quad \mu_i = \sqrt{\sum_{j=1}^{p} \|L_{ji}\|^2}, \ i = 1, \ldots, m, \text{ and } \mu_{m+j} = \sqrt{\sum_{i=1}^{m} \|L_{ji}\|^2}, \ j = 1, \ldots, p,$$

let $\varepsilon \in]0, 1/(\beta + 1)[$ and $(\gamma_n)_{n \geq 0}$ be a sequence in $[\varepsilon, (1 - \varepsilon)/\beta]$. Let the initial points $(x_{1,1,0}, \ldots, x_{1,m,0}) \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_m$, $(x_{2,1,0}, \ldots, x_{2,m,0}) \in \mathcal{H}_1' \times \ldots \times \mathcal{H}_m'$ and $(v_{1,1,0}, \ldots, v_{1,p,0})$, $(v_{2,1,0}, \ldots, v_{2,p,0}) \in \mathcal{G}_1 \times \ldots \times \mathcal{G}_p$ be arbitrary chosen and set

$$
(\forall n \geq 0) \quad \left\lfloor \begin{array}{l}
\text{For } i = 1, \ldots, m \\
\left\lfloor \begin{array}{l}
y_{1,i,n} = x_{1,i,n} - \gamma_n \left(K_i^* x_{2,i,n} + \sum_{j=1}^{p} L_{ji}^* v_{1,j,n} + a_{1,i,n} \right) \\
y_{2,i,n} = x_{2,i,n} - \gamma_n (\nabla h_i^* x_{2,i,n} - K_i x_{1,i,n} + a_{2,i,n}) \\
p_{1,i,n} = y_{1,i,n} + b_{1,i,n} \\
p_{2,i,n} = \text{Prox}_{\gamma_n f_i^*} y_{2,i,n} + b_{2,i,n}
\end{array} \right. \\
\text{For } j = 1, \ldots, p \\
\left\lfloor \begin{array}{l}
w_{1,j,n} = v_{1,j,n} - \gamma_n \left(v_{2,j,n} - \sum_{i=1}^{m} L_{ji} x_{1,i,n} + a_{1,m+j,n} \right) \\
w_{2,j,n} = v_{2,j,n} - \gamma_n (-v_{1,j,n} + a_{2,m+j,n}) \\
r_{1,j,n} = w_{1,j,n} + b_{1,m+j,n} \\
r_{2,j,n} = \text{Prox}_{\gamma_n g_j} w_{2,j,n} + b_{2,m+j,n}
\end{array} \right. \\
\text{For } i = 1, \ldots, m \\
\left\lfloor \begin{array}{l}
q_{1,i,n} = p_{1,i,n} - \gamma_n \left(K_i^* p_{2,i,n} + \sum_{j=1}^{p} L_{ji}^* r_{1,j,n} + c_{1,i,n} \right) \\
q_{2,i,n} = p_{2,i,n} - \gamma_n (\nabla h_i^* p_{2,i,n} - K_i p_{1,i,n} + c_{2,i,n}) \\
x_{1,i,n+1} = x_{1,i,n} - y_{1,i,n} + q_{1,i,n} \\
x_{2,i,n+1} = x_{2,i,n} - y_{2,i,n} + q_{2,i,n}
\end{array} \right. \\
\text{For } j = 1, \ldots, p \\
\left\lfloor \begin{array}{l}
s_{1,j,n} = r_{1,j,n} - \gamma_n \left(r_{2,j,n} - \sum_{i=1}^{m} L_{ji} p_{1,i,n} + c_{1,m+j,n} \right) \\
s_{2,j,n} = r_{2,j,n} - \gamma_n (-r_{1,j,n} + c_{2,m+j,n}) \\
v_{1,j,n+1} = v_{1,j,n} - w_{1,j,n} + s_{1,j,n} \\
v_{2,j,n+1} = v_{2,j,n} - w_{2,j,n} + s_{2,j,n}.
\end{array} \right.
\end{array} \right.
$$

The following convergence result for Algorithm 4.2 is a consequence of Theorem 3.1.

Theorem 4.1. *Suppose that the optimization problem* (4.1) *has an optimal solution and that one of the qualification conditions* (QC_i), $i = 1, 2$, *is fulfilled. For the sequences generated by Algorithm 4.2 the following statements are true:*

(i) $(\forall i \in \{1, \ldots, m\})$ $\sum_{n \geq 0} \|x_{1,i,n} - p_{1,i,n}\|_{\mathcal{H}_i}^2 < +\infty$, $\sum_{n \geq 0} \|x_{2,i,n} - p_{2,i,n}\|_{\mathcal{H}_i}^2 < +\infty$ *and* $(\forall j \in \{1, \ldots, p\})$ $\sum_{n \geq 0} \|v_{1,j,n} - r_{1,j,n}\|_{\mathcal{G}_j}^2 < +\infty$, $\sum_{n \geq 0} \|v_{2,j,n} - r_{2,j,n}\|_{\mathcal{G}_j}^2 < +\infty$.

(ii) *There exists an optimal solution* $(\overline{x}_1, \ldots, \overline{x}_m)$ *to* (4.1) *and an optimal solution* $(\overline{w}_1, \ldots, \overline{w}_m, \overline{w}_{m+1}, \ldots, \overline{w}_{m+p})$ *to* (4.8), *such that* $(\forall i \in \{1, \ldots, m\})$ $x_{1,i,n} \rightharpoonup \overline{x}_i$, $p_{1,i,n} \rightharpoonup \overline{x}_i$, $x_{2,i,n} \rightharpoonup \overline{w}_i$ *and* $p_{2,i,n} \rightharpoonup \overline{w}_i$ *and* $(\forall j \in \{1, \ldots, p\})$ $v_{1,j,n} \rightharpoonup \overline{w}_{m+j}$ *and* $r_{1,j,n} \rightharpoonup \overline{w}_{m+j}$ *as* $n \to +\infty$.

Remark 4.4. Recently, in [16], another iterative scheme for solving systems of monotone inclusions, that is also able to handle with the solving of optimization problems of type (4.1), in case when the functions $g_j, j = 1, ..., p$, are not necessarily differentiable, was proposed. Different to our approach, which assumes that the variables are coupled by the single-valued operator B, in [16] the coupling is made by some compositions of parallel sums of maximally monotone operators with linear continuous ones.

5. Numerical Experiments

In this section we present two numerical experiments which emphasize the performances of the primal-dual algorithm for systems of monotone inclusions.

5.1. Average consensus for colored networks

The first numerical experiment that we consider concerns the problem of average consensus on colored networks.

Given a network, where each node posses a measurement in form of a real number, the average consensus problem consists in calculating the average of these measurements in a recursive and distributed way, allowing the nodes to communicate information along only the available edges in the network. Consider a connected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of *nodes* and $\mathcal{E}$ represents the set of *edges*. Each edge is uniquely represented as a pair of nodes (i, j), where $i < j$. The nodes i and j can exchange their values if they can communicate directly, in other words, if $(i, j) \in \mathcal{E}$. We assume that each node possesses a measurement in form of a real number, also called *color*, and that no neighboring nodes have the same color. Let C denote the number of colors the network is colored with and $\mathcal{C}_i$ the set of the nodes that have the color $i,\ i = 1, \ldots, C$. Without affecting the generality of the problem we also assume that the first C_1 nodes are in the set $\mathcal{C}_1$, the next C_2 nodes are in the set $\mathcal{C}_2$, etc. Furthermore, we assume that a node coloring scheme is available. For more details concerning the mathematical modelling of the average consensus problem on colored networks we refer the reader to [22, 23].

Let P and E denote the number of nodes and edges in the network, respectively, hence, $\sum_{i=1}^{C} C_i = P$. Denoting by θ_k the measurement assigned to node k, $k = 1, \ldots, P$, the problem we want to solve is

$$\min_{x \in \mathbb{R}} \sum_{k=1}^{P} \frac{1}{2}(x - \theta_k)^2. \tag{5.1}$$

The unique optimal solution to the problem (5.1) is $\theta^* = \frac{1}{P}\sum_{k=1}^{P} \theta_k$, namely the average of the measurements over the whole set of nodes in the network. The goal is to make this value available in each node in a distributed and recursive way. To this end, we replicate copies of x throughout the entire network, more precisely, for $k = 1, ..., P$, node k will hold the k-th copy, denoted by x_k, which will be updated iteratively during the algorithm. At the end we have to guarantee that all the copies are equal and we express this constraint by requiring that $x_i = x_j$ for each $(i,j) \in \mathcal{E}$. This gives rise to the following optimization problem

$$\min_{\substack{\overline{x}=(x_1,\ldots,x_P)\in\mathbb{R}^P \\ x_i=x_j,\ \forall\{i,j\}\in\mathcal{E}}} \sum_{k=1}^{P} \frac{1}{2}(x_k-\theta_k)^2. \tag{5.2}$$

Let $A \in \mathbb{R}^{P\times E}$ be the *node-arc incidence matrix* of the network, which is the matrix having each column associated to an edge in the following manner: the column associated to the edge $(i,j) \in \mathcal{E}$ has 1 at the i-th entry and -1 at the j-th entry, the remaining entries being equal to zero. Consequently, constraints in (5.2) can be written with the help of the *transpose* of the node-arc incidence matrix as $A^T\overline{x} = 0$. Taking into consideration the ordering of the nodes and the coloring scheme, we can write $A^T\overline{x} = A_1^T\overline{x}_1 + \ldots + A_C^T\overline{x}_C$, where $\overline{x}_i \in \mathbb{R}^{C_i}$, $i = 1,...,C$, collects the copies of the nodes in $\mathcal{C}_i$, i.e.

$$\overline{x} = (\underbrace{x_1,...,x_{C_1}}_{\overline{x}_1},...,\underbrace{x_{P-C_C+1},...,x_P}_{\overline{x}_C}).$$

Hence, the optimization problem (5.2) becomes

$$\min_{\substack{\overline{x}=(\overline{x}_1,\ldots,\overline{x}_C) \\ A_1^T\overline{x}_1+\ldots+A_C^T\overline{x}_C=0}} \sum_{i=1}^{C} f_i(\overline{x}_i), \tag{5.3}$$

where for $i = 1,...,C$, the function $f_i : \mathbb{R}^{C_i} \to \mathbb{R}$ is defined as $f_i(\overline{x}_i) = \sum_{l\in\mathcal{C}_i} \frac{1}{2}(x_l - \theta_l)^2$.

One can easily observe that problem (5.3) is a particular instance of the optimization problem (4.1), when taking

$$m = C, p = 1, h_i = \delta_{\{0\}}, K_i = \mathrm{Id}\ , L_{1i} = A_i^T \in \mathbb{R}^{E\times C_i}, i = 1,...,C, \text{ and } g_1 = \delta_{\{0\}}.$$

Using that $h_i^* = 0, i = 1,...,C$, and $\mathrm{Prox}_{\gamma g}(x) = 0$ for all $\gamma > 0$ and $x \in \mathbb{R}^E$, the iterative scheme in Algorithm 4.2 reads, after some algebraic manipulations, in the error-free case:

$$(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} \text{For } i = 1,\ldots,C \\ \quad \left\lfloor \begin{array}{l} y_{1,i,n} = x_{1,i,n} - \gamma_n\big(x_{2,i,n} + A_i v_{1,1,n}\big) \\ y_{2,i,n} = x_{2,i,n} + \gamma_n x_{1,i,n} \\ p_{2,i,n} = \mathrm{Prox}_{\gamma_n f_i^*} y_{2,i,n} \end{array}\right. \\ w_{1,1,n} = v_{1,1,n} - \gamma_n\big(v_{2,1,n} - \sum_{i=1}^C A_i^T x_{1,i,n}\big) \\ \text{For } i = 1,\ldots,C \\ \quad \left\lfloor \begin{array}{l} q_{1,i,n} = y_{1,i,n} - \gamma_n\big(p_{2,i,n} + A_i w_{1,1,n}\big) \\ q_{2,i,n} = p_{2,i,n} + \gamma_n y_{1,i,n} \\ x_{1,i,n+1} = x_{1,i,n} - y_{1,i,n} + q_{1,i,n} \\ x_{2,i,n+1} = x_{2,i,n} - y_{2,i,n} + q_{2,i,n} \end{array}\right. \\ v_{1,1,n+1} = v_{1,1,n} + \gamma_n \sum_{i=1}^C A_i^T y_{1,i,n} \\ v_{2,1,n+1} = \gamma_n^2\big(\sum_{i=1}^C A_i^T x_{1,i,n} - v_{2,1,n}\big). \end{array}\right.$$

Let us notice that for $i = 1, ..., C$ and $\gamma > 0$ it holds $\text{Prox}_{\gamma f_i^*}(\overline{x}_i) = (1+\gamma)^{-1}(\overline{x}_i - \gamma\overline{\theta}_i)$, where $\overline{\theta}_i$ is the vector in $\mathbb{R}^{C_i}$ whose components are θ_l with $l \in \mathcal{C}_i$. In order to compare the performances of our method with other existing algorithms in literature, we used the networks generated in [23] with the number of nodes ranging between 10 and 1000. The measurement θ_k associated to each node was generated randomly and independently from a normal distribution with mean 10 and standard deviation 100. We worked with networks with 10, 50, 100, 200, 500, 700 and 1000 nodes and measured the performance of our algorithm from the point of view of the number of communication steps, which actually coincides with the number of iterations. As stopping criterion we considered

$$\frac{\|x_n - \mathbb{1}_P\theta^*\|}{\left|\sqrt{P}\theta^*\right|} \leq 10^{-4},$$

where $\mathbb{1}_P$ denotes the vector in $\mathbb{R}^P$ having all entries equal to 1.

(a) Watts-Strogatz network (b) Geometric network

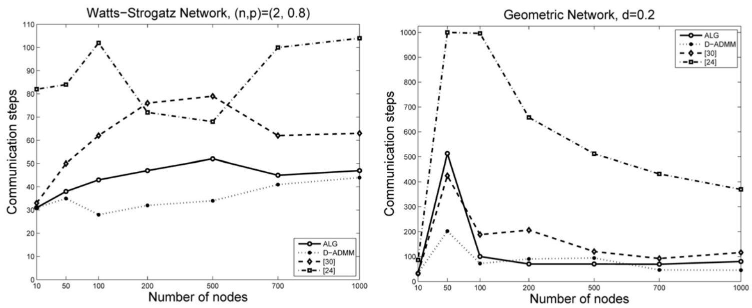

Fig. 5.1. Figure (a) shows the communication steps needed by the four algorithms for a Watts-Strogatz network with different number of nodes. Figure (b) shows the communication steps needed by the four algorithms for a Geometric network with different number of nodes. In both figures ALG stands for the primal-dual algorithm proposed in this paper.

We present in Figure 5.1 the communication steps needed when dealing with the *Watts-Strogatz network* with parameters (2,0.8) and with the *Geometric network* with a distance parameter 0.2. The Watts-Strogatz network is created from a lattice where every node is connected to 2 nodes, then the links are rewired with a probability of 0.8, while the Geometric network works with nodes in a $[0, 1]^2$ square and connects the nodes whose Euclidean distance is less than the given parameter 0.2. As shown in Figure 5.1, our algorithm performed comparable to D-AMM, presented in [23], but it performed better then the algorithms presented in [24] and [31].

In order to observe the behavior of our algorithm on different networks, we tested it on the following 6 networks: 1. Erdős-Rényi network with parameter 0.25, 2. Watts-Strogatz network with parameters (2, 0.8) (network 2), 3. Watts-Strogatz network with parameters (4, 0.6), 4. Barabási-Albert network, 5. Geometric network with parameter 0.2 and 6. Lattice network, with a different number of nodes. Observing the needed communication steps, we can conclude that our algorithm is communication-efficient and it performs better than or similarly to the algorithms in [23, 24] and [31] (as exemplified in Figure 5.2).

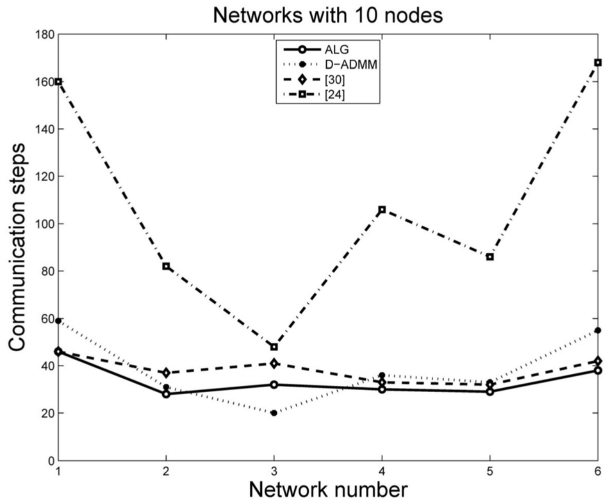

Fig. 5.2. Comparison of the four algorithms over six networks with 10 nodes. Here, ALG stands for the primal-dual algorithm proposed in this paper.

5.2. Support vector machines classification

The second numerical experiment we present in this section addresses the problem of classifying images via support vector machines.

Having a set training data $a_i \in \mathbb{R}^n,\ i = 1,\dots,k$, belonging to one of two given classes, denoted by "-1" and "+1", the aim is to construct by it a decision function given in the form of a separating hyperplane which should assign every new data to one of the two classes with a low misclassification rate. We construct the matrix $A \in \mathbb{R}^{k\times n}$ such that each row corresponds to a data point $a_i, i = 1, ..., k$ and a vector $d \in \mathbb{R}^k$ such that for $i = 1, ..., k$ its i-th entry is equal to -1, if a_i belongs to the class "-1" and it is equal to $+1$, otherwise. In order to cover the situation when the separation cannot be done exactly, we consider non-negative slack variables $\xi_i \geq 0,\ i = 1,\dots,k$, thus the goal will be to find $(s, r, \xi) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^k_+$ as optimal solution of the following optimization problem (also called *soft-margin support vector machines problem*)

$$\min_{\substack{(s,r,\xi)\in\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}^k_+\\ D(As+1_k r)+\xi\geqq 1_k}} \left\{\|s\|^2 + C\,\|\xi\|^2\right\}, \tag{5.4}$$

where $\mathbb{1}_k$ denotes the vector in $\mathbb{R}^k$ having all entries equal to 1, the inequality $z \geqq \mathbb{1}_k$ for $z \in \mathbb{R}^k$ means $z_i \geq 1, i = 1, ...k$, $D = \text{Diag}(d)$ is the diagonal matrix having the vector d as main diagonal and C is a trade-off parameter. Each new data $a \in \mathbb{R}^n$ will by assigned to one of the two classes by means of the resulting decision function $z(a) = s^T a + r$, namely, a will be assigned to the class "-1", if $z(a) < 0$, and to the class "+1", otherwise. For more theoretical insights in support vector machines we refer the reader to [19].

A sample of data for number 2 A sample of data for number 9

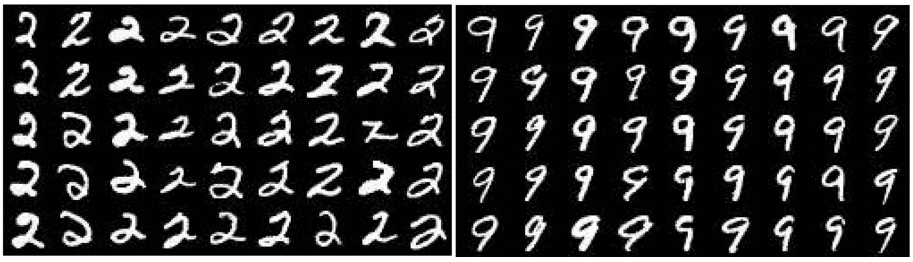

Fig. 5.3. A sample of images belonging to the classes -1 and +1, respectively.

We made use of a data set of 11907 training images and 2041 test images of size 28×28 from the website `http://www.cs.nyu.edu/~roweis/data.html`. The problem consisted in determining a decision function based on a pool of handwritten digits showing either the number two or the number nine, labeled by -1 and $+1$, respectively (see Figure 5.3). We evaluated the quality of the decision function on a test data set by computing the percentage of misclassified images. Notice that we use only a half of the available images from the training data set, in order to reduce the computational effort.

The soft-margin support vector machines problem (5.4) can be written as a special instance of the optimization problem (4.1), by taking

$$m = 3, p = 1, f_1(\cdot) = \|\cdot\|^2, f_2 = 0, f_3(\cdot) = C\,\|\cdot\|^2 + \delta_{\mathbb{R}^k_+}(\cdot),$$

$$h_i = \delta_{\{0\}}, K_i = \text{Id}\ , i = 1, 2, 3,$$

$$g_1 = \delta_{\{z\in\mathbb{R}^k : z\geqq 1_k\}}, L_{11} = DA, L_{12} = D1_k \text{ and } L_{13} = \text{Id}\ .$$

Thus, Algorithm 4.2 gives rise in the error-free case to the following iterative scheme:

$$
(\forall n \geq 0)\quad \left\lfloor \begin{array}{l}
\text{For } i = 1,2,3 \\
\quad \left\lfloor \begin{array}{l}
y_{1,i,n} = x_{1,i,n} - \gamma_n\big(x_{2,i,n} + L_{1i}^T v_{1,1,n}\big) \\
y_{2,i,n} = x_{2,i,n} + \gamma_n x_{1,i,n} \\
p_{2,i,n} = \text{Prox}_{\,\gamma_n f_i^*} y_{2,i,n}
\end{array}\right. \\
w_{1,1,n} = v_{1,1,n} - \gamma_n\big(v_{2,1,n} - \sum_{i=1}^3 L_{1i} x_{1,i,n}\big) \\
w_{2,1,n} = v_{2,1,n} + \gamma_n v_{1,1,n} \\
r_{2,1,n} = \text{Prox}_{\,\gamma_n g_1} w_{2,1,n} \\
\text{For } i = 1,2,3 \\
\quad \left\lfloor \begin{array}{l}
q_{1,i,n} = y_{1,i,n} - \gamma_n\big(p_{2,i,n} + L_{1i}^T w_{1,1,n}\big) \\
q_{2,i,n} = p_{2,i,n} + \gamma_n y_{1,i,n} \\
x_{1,i,n+1} = x_{1,i,n} - y_{1,i,n} + q_{1,i,n} \\
x_{2,i,n+1} = x_{2,i,n} - y_{2,i,n} + q_{2,i,n}
\end{array}\right. \\
s_{1,1,n} = w_{1,1,n} - \gamma_n(r_{2,1,n} - \sum_{i=1}^3 L_{1i} p_{1,i,n}) \\
s_{2,1,n} = r_{2,1,n} + \gamma_n w_{1,1,n} \\
v_{1,1,n+1} = v_{1,1,n} - w_{1,1,n} + s_{1,1,n} \\
v_{2,1,n+1} = v_{2,1,n} - w_{2,1,n} + s_{2,1,n}
\end{array}\right.
$$

We would also like to notice that for the proximal points needed in the algorithm one has for $\gamma > 0$ and $(s, r, \xi, z) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^k$ the following exact formulae:

$$\text{Prox}_{\,\gamma f_1^*}(s) = (2+\gamma)^{-1}2s, \text{Prox}_{\,\gamma f_2^*}(r) = 0, \text{Prox}_{\,\gamma f_3^*}(\xi) = \xi - \gamma P_{\mathbb{R}_+^k}\big((2C+\gamma)^{-1}\xi\big)$$

and

$$\text{Prox}_{\,\gamma g_1}(z) = P_{\{x\in\mathbb{R}^k : x \geqq 1_k\}}(z).$$

With respect to the considered data set, we denote by $\mathcal{D} = \{(X_i, Y_i), i = 1, ..., 6000\} \subseteq \mathbb{R}^{784} \times \{+1, -1\}$ the set of available training data consisting of 3000 images in the class -1 and 3000 images in the class $+1$. A sample from each class of images is shown in Figure 5.3. The images have been vectorized and normalized by dividing each of them by the quantity $\left(\frac{1}{6000}\sum_{i=1}^{6000} \|X_i\|^2\right)^{\frac{1}{2}}$. We stopped the primal-dual algorithm after different numbers of iterations and evaluated the performances of the resulting decision functions. In Table 1 we present the misclassification rate in percentage for the training and for the test data (the error for the training data is less than the one for the test data) and observe that the quality of the classification increases with the number of iterations. However, even for a low number of iterations the misclassification rate outperforms the ones reported in the literature dealing with numerical methods for support vector classification. Let us also mention that the numerical results are given for the case $C = 1$. We tested also other choices for C, however we did not observe great impact on the results.

Table 1. Misclassification rate in percentage for different numbers of iterations for both the training data and the test data.

Number of iterations	100	1000	2000	3000	5000
Training error	2.95	2.6	2.3	1.95	1.55
Test error	2.95	2.55	2.45	2.15	2

Acknowledgments

The authors are grateful to Christopher Hendrich for helpful hints concerning the implementation issues of the algorithm, to João F. C. Mota for sharing the data he worked with and to an anonoymous reviewer for remarks which improved the quality of the presentation.

References

1. H. Attouch, L. M. Briceño-Arias and P. L. Combettes, A parallel splitting method for coupled monotone inclusions, *SIAM Journal on Control and Optimization*, **48(5)** (2010), 3246-3270.
2. H. Attouch and M. Théra, A general duality principle for the sum of two operators, *Journal of Convex Analysis*, **3** (1996), 1-24.
3. H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer-Verlag, New York, 2011.
4. H. H. Bauschke, P. L. Combettes and S. Reich, The asymptotic behavior of the composition of two resolvents, *Nonlinear Analysis*, **60(2)** (2005), 283-301.
5. R. I. Boţ, *Conjugate Duality in Convex Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 637, Springer-Verlag, Berlin Heidelberg, 2010.
6. R. I. Boţ and E. R. Csetnek, Regularity conditions via generalized interiority notions in convex optimization: New achievements and their relation to some classical statements, *Optimization*, **61(1)** (2012), 35-65.
7. R. I. Boţ, E. R. Csetnek and A. Heinrich, *A primal-dual splitting algorithm for finding zeros of sums of maximally monotone operators*, arXiv: 1206.5953v1 [math.OC], 2012.
8. R. I. Boţ and C. Hendrich, *A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, arXiv: 1212.0326v1 [math.OC], 2012.
9. R. I. Boţ and C. Hendrich, *Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization*, arXiv: 1211.1706v1 [math.OC], 2012.

10. R. I. Boţ and C. Hendrich, On the acceleration of the double smoothing technique for unconstrained convex optimization problems, *Optimization*, 2012, http://dx.doi.org/10.1080/02331934.2012.745530.

11. L. M. Briceño-Arias and P. L. Combettes, A monotone + skew splitting model for composite monotone inclusions in duality, *SIAM Journal on Optimization*, **21(4)** (2011), 1230-1250.

12. P. Calamai and C. Charalambous, Solving multifacility location problems involving Euclidean distances, *Naval Research Logistics Quarterly*, **27(4)** (1980), 609-620.

13. A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *Journal of Mathematical Imaging and Vision*, **40(1)** (2011), 120-145.

14. P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization*, **53(5-6)** (2004), 475-504.

15. P. L. Combettes, Iterative construction of the resolvent of a sum of maximal monotone operators, *Journal of Convex Analysis*, **16(3-4)** (2009), 727-748.

16. P. L. Combettes, *Systems of structured monotone inclusions: duality, algorithms, and applications*, arXiv: 1212.6631v1 [math.OC], 2012.

17. 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 and Variational Analysis*, **20(2)** (2012), 307-330.

18. L. Condat, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, *Journal of Optimization Theory and Applications*, DOI 10.1007/s10957-012-0245-9, 2012.

19. N. Cristianini and J. S. Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, 2000.

20. I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland Publishing Company, Amsterdam, 1976.

21. J. W. Eyster, J. A. White and W. W. Wierwille, On solving multifacility location problems using a hyperboloid approximation procedure, *AIIE Transactions*, **5(1)** (1973), 1-6.

22. J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar and M. Püschel, *ADMM for consensus on colored networks*, 2012 IEEE 51st Annual Conference on Decision and Control (CDC), 2012, pp. 5116-5121.

23. J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar and M. Püschel, *D-ADMM: A communication-efficient distributed algorithm for separable optimization*, arXiv: 1202.2805 [math.OC], 2012.

24. A. Olshevsky and J. Tsitsiklis, Convergence speed in distributed consensus and averaging, *SIAM Journal on Control and Optimization*, **48(1)** (2009), 33-55.

25. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

26. T. Strömberg, *The Operation of Infimal Convolution*, Institute of Mathematics, Polish Academy of Sciences, Warsaw, 1996.

27. S. Simons, *From Hahn-Banach to Monotonicity*, Springer-Verlag, Berlin, 2008.

28. P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM Journal on Control and Optimization*, **38(2)** (2000), 431-446.

29. B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Advances in Computational Mathematics*, **38(3)** (2013), 667-681.

30. C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, Singapore, 2002.

31. H. Zhu, G. Giannakis and A. Cano, Distributed in-network channel decoding, *IEEE Transactions on Signal Processing*, **57(10)** (2009), 3970-3983.

Radu Ioan Boţ and Ernö Robert Csetnek
Department of Mathematics
Chemnitz University of Technology
D-09107 Chemnitz
Germany
E-mail: radu.bot@mathematik.tu-chemnitz.de
robert.csetnek@mathematik.tu-chemnitz.de

Erika Nagy
Faculty of Mathematics and Computer Science
Babeş-Bolyai University
Cluj-Napoca, Romania
E-mail: erika.nagy@ubbcluj.ro