On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems

Radu Ioan Boţ [*] Ernö Robert Csetnek [†] André Heinrich [‡]
Christopher Hendrich [§]

February 10, 2014

Abstract. We present two modified versions of the primal-dual splitting algorithm relying on forward-backward splitting proposed in [27] for solving monotone inclusion problems. Under strong monotonicity assumptions for some of the operators involved we obtain for the sequences of iterates that approach the solution orders of convergence of $\mathcal{O}(\frac{1}{n})$ and $\mathcal{O}(\omega^n)$, for $\omega \in (0,1)$, respectively. The investigated primal-dual algorithms are fully decomposable, in the sense that the operators are processed individually at each iteration. We also discuss the modified algorithms in the context of convex optimization problems and present numerical experiments in image processing and pattern recognition in cluster analysis.

Key Words. maximally monotone operator, strongly monotone operator, resolvent, operator splitting, subdifferential, strongly convex function, convex optimization algorithm, duality

AMS subject classification. 47H05, 65K05, 90C25

1 Introduction and preliminaries

The problem of finding the zeros of the sum of two (or more) maximally monotone operators in Hilbert spaces continues to be a very active research field, with applications in convex optimization, partial differential equations, signal and image processing, etc. (see [1, 6–9, 13, 14, 27]). To the most prominent methods in this area belong the proximal point algorithm for finding the zeros of a maximally monotone operator (see [23]) and the Douglas-Rachford splitting algorithm for finding the zeros of the sum of two maximally

[*]University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: radu.bot@univie.ac.at. Research partially supported by DFG (German Research Foundation), project BO 2516/4-1.

[†]University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: ernoe.robert.csetnek@univie.ac.at. Research supported by DFG (German Research Foundation), project BO 2516/4-1.

[‡]Chemnitz University of Technology, Department of Mathematics, D-09107 Chemnitz, Germany, e-mail: andre.heinrich@mathematik.tu-chemnitz.de. Research supported by the European Union, the European Social Fund (ESF) and prudsys AG in Chemnitz.

[§]Chemnitz University of Technology, Department of Mathematics, D-09107 Chemnitz, Germany, e-mail: christopher.hendrich@mathematik.tu-chemnitz.de. Research supported by a Graduate Fellowship of the Free State Saxony, Germany.

monotone operators (see [16]). However, also motivated by different applications, the research community was interested in considering more general problems, in which the sum of finitely many operators appear, some of them being composed with linear continuous operators [1, 9, 13]. In the last years, even more complex structures were considered, in which also parallel sums are involved, see [7, 8, 14, 27].

The algorithms introduced in the literature for these issues have the remarkable property that the operators involved are evaluated separately in each iteration, either by forward steps in the case of the single-valued ones (including here the linear continuous operators and their adjoints) or by backward steps for the set-valued ones, by using the corresponding resolvents. More than that they share the common feature to be of primal-dual type, meaning that they solve not only the primal inclusion problem, but also its Attouch-Théra-type dual. In this context we mention the primal-dual algorithms relying on Tseng's forward-backward-forward splitting method (see [9, 14]), on the forward-backward splitting method (see [27]) and on the Douglas-Rachford splitting method (see [8]). A relevant task is to adapt these iterative methods in order be able to investigate their convergence, namely, to eventually determine convergence rates for the sequences generated by the schemes in discussion. This could be important when one is interested in obtaining an optimal solution more rapidly than in their initial formulation, which furnish "only" the convergence statement. Accelerated versions of the primal-dual algorithm from [14] were already provided in [7], whereby the reported numerical experiments emphasize the advantages of the first over the original iterative scheme.

The aim of this paper is to provide modified versions of the algorithm proposed by Vũ in [27] for which an evaluation of their convergence behaviour is possible. By assuming that some of the operators involved are strongly monotone, we are able to obtain for the sequences of iterates orders of convergence of $\mathcal{O}(\frac{1}{n})$ and $\mathcal{O}(\omega^n)$, for $\omega \in (0,1)$, respectively.

For the readers convenience we present first some notations which are used throughout the paper (see [1, 3, 4, 17, 24, 28]). Let $\mathcal{H}$ be a real Hilbert space with *inner product* $\langle \cdot, \cdot \rangle$ and associated *norm* $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. The symbols $\rightharpoonup$ and $\to$ denote weak and strong convergence, respectively. When $\mathcal{G}$ is another Hilbert space and $K : \mathcal{H} \to \mathcal{G}$ a linear continuous operator, then the *norm* of K is defined as $\|K\| = \sup\{\|Kx\| : x \in \mathcal{H}, \|x\| \leq 1\}$, while $K^* : \mathcal{G} \to \mathcal{H}$, defined by $\langle K^*y, x\rangle = \langle y, Kx\rangle$ for all $(x,y) \in \mathcal{H} \times \mathcal{G}$, denotes the *adjoint operator* of K.

For an arbitrary set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ we denote by $\operatorname{Gr} A = \{(x,u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ its *graph*, by $\operatorname{dom} A = \{x \in \mathcal{H} : Ax \neq \emptyset\}$ its *domain* and by $A^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ its *inverse operator*, defined by $(u,x) \in \operatorname{Gr} A^{-1}$ if and only if $(x,u) \in \operatorname{Gr} A$. We say that A is *monotone* if $\langle x - y, u - v\rangle \geq 0$ for all $(x,u), (y,v) \in \operatorname{Gr} A$. A monotone operator A is said to be *maximally monotone*, if there exists no proper monotone extension of the graph of A on $\mathcal{H} \times \mathcal{H}$. The *resolvent* of A, $J_A : \mathcal{H} \rightrightarrows \mathcal{H}$, is defined by $J_A = (\operatorname{Id}_{\mathcal{H}} + A)^{-1}$, where $\operatorname{Id}_{\mathcal{H}} : \mathcal{H} \to \mathcal{H}, \operatorname{Id}_{\mathcal{H}}(x) = x$ for all $x \in \mathcal{H}$, is the *identity operator* on $\mathcal{H}$. Moreover, if A is maximally monotone, then $J_A : \mathcal{H} \to \mathcal{H}$ is single-valued and maximally monotone (cf. [1, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [1, Proposition 23.2])

$$p \in J_{\gamma A}x \text{ if and only if } (p, \gamma^{-1}(x - p)) \in \operatorname{Gr} A$$

and (see [1, Proposition 23.18])

$$J_{\gamma A} + \gamma J_{\gamma^{-1}A^{-1}} \circ \gamma^{-1} \operatorname{Id}_{\mathcal{H}} = \operatorname{Id}_{\mathcal{H}}. \tag{1}$$

Let $\gamma > 0$ be arbitrary. We say that A is *γ-strongly monotone* if $\langle x-y, u-v\rangle \geq \gamma\|x-y\|^2$ for all $(x,u),(y,v) \in \operatorname{Gr} A$. A single-valued operator $A : \mathcal{H} \to \mathcal{H}$ is said to be *γ-cocoercive* if $\langle x-y, Ax-Ay\rangle \geq \gamma\|Ax-Ay\|^2$ for all $(x,y) \in \mathcal{H}\times\mathcal{H}$. Moreover, A is *γ-Lipschitzian* if $\|Ax-Ay\| \leq \gamma\|x-y\|$ for all $(x,y) \in \mathcal{H}\times\mathcal{H}$. A single-valued linear operator $A : \mathcal{H} \to \mathcal{H}$ is said to be *skew*, if $\langle x, Ax\rangle = 0$ for all $x \in \mathcal{H}$. Finally, the *parallel sum* of two operators $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined by $A\Box B : \mathcal{H} \rightrightarrows \mathcal{H}, A\Box B = (A^{-1}+B^{-1})^{-1}$.

The following problem represents the starting point of our investigations (see [27]).

Problem 1 Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ an η-cocoercive operator for $\eta > 0$. Let m be a strictly positive integer and, for any $i \in \{1,...,m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone operator, let $D_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone and ν_i-strongly monotone operator for $\nu_i > 0$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^*\big((B_i\Box D_i)(L_i\overline{x}-r_i)\big) + C\overline{x}, \tag{2}$$

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \begin{cases} z - \sum_{i=1}^m L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in (B_i\Box D_i)(L_ix - r_i),\ i = 1,...,m. \end{cases} \tag{3}$$

We say that $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H}\times\ \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to Problem 1, if

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in A\overline{x} + C\overline{x} \text{ and } \overline{v}_i \in (B_i\Box D_i)(L_i\overline{x}-r_i),\ i = 1,...,m. \tag{4}$$

If $\overline{x} \in \mathcal{H}$ is a solution to (2), then there exists $(\overline{v}_1,..., \overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$ such that $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ is a primal-dual solution to Problem 1 and, if $(\overline{v}_1,..., \overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a solution to (3), then there exists $\overline{x} \in \mathcal{H}$ such that $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ is a primal-dual solution to Problem 1. Moreover, if $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H}\times\ \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to Problem 1, then $\overline{x}$ is a solution to (2) and $(\overline{v}_1,..., \overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a solution to (3).

By employing the classical forward-backward algorithm (see [13, 26]) in a renormed product space, Vũ proposed in [27] an iterative scheme for solving a slightly modified version of Problem 1 formulated in the presence of some given weights $w_i \in (0,1]$, $i = 1,...,m$, with $\sum_{i=1}^m w_i = 1$ for the terms occurring in the second summand of the primal inclusion problem. The following result is an adaption of [27, Theorem 3.1] to Problem 1 in the error-free case and when $\lambda_n = 1$ for any $n \geq 0$.

Theorem 2 *In Problem 1 suppose that*

$$z \in \operatorname{ran}\left(A + \sum_{i=1}^{m} L_i^*\big((B_i\Box D_i)(L_i\cdot - r_i)\big) + C\right).$$

Let τ and σ_i, $i = 1,...,m$, be strictly positive numbers such that

$$2\cdot\min\{\tau^{-1}, \sigma_1^{-1},..., \sigma_m^{-1}\}\cdot\min\{\eta, \nu_1,..., \nu_m\}\left(1 - \sqrt{\tau\sum_{i=1}^{m}\sigma_i\|L_i\|^2}\right) > 1.$$

Let $(x_0, v_{1,0},..., v_{m,0}) \in \mathcal{H}\times \mathcal{G}_1 \times ...\times \mathcal{G}_m$ and for all $n \geq 0$ set:

$$\begin{aligned}
&x_{n+1} = J_{\tau A}\big[x_n - \tau\big(\textstyle\sum_{i=1}^m L_i^* v_{i,n} + Cx_n - z\big)\big]\\
&y_n = 2x_{n+1} - x_n\\
&v_{i,n+1} = J_{\sigma_i B_i^{-1}}[v_{i,n} + \sigma_i(L_i y_n - D_i^{-1} v_{i,n} - r_i)],\ i = 1,..., m.
\end{aligned}$$

Then there exists a primal-dual solution $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ to Problem 1 such that $x_n \rightharpoonup \overline{x}$ and $(v_{1,n},..., v_{m,n}) \rightharpoonup (\overline{v}_1,..., \overline{v}_m)$ as $n \to +\infty$.

Notice that the work in [27] is closely related to [11] and [15], where primal-dual splitting methods for nonsmooth convex optimization problems are proposed. More exactly, the convergence property of [11, Algorithm 1] provided in [11, Theorem 1] follow as special instance of the main result in [27]. On the other hand, Condat proposes in [15] an algorithm which can be seen as an extension of the one in [11] to optimization problems in the objective of which convex differentiable functions occurr, as well.

The structure of the paper is as follows. In the next section we propose under appropriate strong monotonicity assumptions two modified versions of the above algorithm which ensure for the sequences of iterates orders of convergence of $\mathcal{O}(\frac{1}{n})$ and $\mathcal{O}(\omega^n)$, for $\omega \in (0, 1)$, respectively. In Section 3 we show how to particularize the general results in the context of nondifferentiable convex optimization problems, where some of the functions occurring in the objective are strongly convex. In the last section we present some numerical experiments in image denoising and pattern recognition in cluster analysis and emphasize also the practical advantages of the modified iterative schemes over the initial one provided in Theorem 2. Numerical comparisons to other state-of-the-art methods for solving convex nondifferentiable optimization problems are also given.

2 Two modified primal-dual algorithms

In this section we propose in two different settings modified versions of the algorithm in Theorem 2 and discuss the orders of convergence of the sequences of iterates generated by the new schemes.

2.1 The case $A + C$ is strongly monotone

For the beginning, we show that, in case $A + C$ is strongly monotone, one can guarantee an order of convergence of $\mathcal{O}(\frac{1}{n})$ for the sequence $(x_n)_{n\geq 0}$. To this end, inspired by [29] and [11], we update in each iteration the parameters τ and σ_i, $i = 1,..., m$, and use a modified formula for the sequence $(y_n)_{n\geq 0}$. Due to technical reasons, we apply this method in case D_i^{-1} is equal to zero for $i = 1,..., m$, that is $D_i(0) = \mathcal{G}_i$ and $D_i(x) = \emptyset$ for $x \neq 0$. Let us notice that, by using the approach proposed in [7, Remark 3.2], one can extend the statement of Theorem 8 below, which is the main result of this subsection, to the primal-dual pair of monotone inclusions as stated in Problem 1.

More precisely, the problem we consider throughout this subsection is as follows.

Problem 3 Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ a monotone and η-Lipschitzian operator for $\eta > 0$. Let m be a strictly positive integer and, for any $i \in \{1,..., m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone operator and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^*(B_i(L_i\overline{x} - r_i)) + C\overline{x}, \tag{5}$$

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,\dots,\overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \begin{cases} z - \sum_{i=1}^m L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in B_i(L_ix - r_i),\ i = 1,\dots,m. \end{cases} \tag{6}$$

As for Problem 1, we say that $(\overline{x}, \overline{v}_1,\dots,\overline{v}_m) \in \mathcal{H}\times\ \mathcal{G}_1 \times \dots \times \mathcal{G}_m$ is a primal-dual solution to Problem 3, if

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in A\overline{x} + C\overline{x} \text{ and } \overline{v}_i \in B_i(L_i\overline{x} - r_i),\ i = 1,\dots,m. \tag{7}$$

Remark 4 One can notice that, in comparison to Problem 1, we relax in Problem 3 the assumptions made on the operator C. It is obvious that, if C is a η-cocoercive operator for $\eta > 0$, then C is monotone and $1/\eta$-Lipschitzian. Although in case C is the gradient of a convex and differentiable function, due to the celebrated Baillon-Haddad Theorem (see, for instance, [1, Corollary 8.16]), the two classes of operators coincide, in general the second one is larger. Indeed, nonzero linear, skew and Lipschitzian operators are not cocoercive. For example, when $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces and $L : \mathcal{H} \to \mathcal{G}$ is nonzero linear continuous, then $(x, v) \mapsto (L^*v, -Lx)$ is an operator having all these properties. This operator appears in a natural way when considering primal-dual monotone inclusion problems as done in [9].

Under the assumption that $A + C$ is γ-strongly monotone for $\gamma > 0$ we propose the following modification of the iterative scheme in Theorem 2.

Algorithm 5

`Initialization`:	`Choose` $\tau_0 > 0, \sigma_{i,0} > 0,\ i = 1,\dots,m,$ `such that`
	$\tau_0 < 2\gamma/\eta,\ \lambda \geq \eta + 1,\ \tau_0 \sum_{i=1}^m \sigma_{i,0}\|L_i\|^2 \leq \sqrt{1 + \tau_0(2\gamma - \eta\tau_0)/\lambda},$
	$\theta_0 = 1/\sqrt{1+\tau_0(2\gamma-\eta\tau_0)/\lambda}$ `and` $(x_0, v_{1,0},\dots, v_{m,0}) \in \mathcal{H}\times\mathcal{G}_1\times\dots\times\mathcal{G}_m$.
`For` $n \geq 0$ `set`:	$x_{n+1} = J_{(\tau_n/\lambda)A}\big[x_n - (\tau_n/\lambda)\big(\sum_{i=1}^m L_i^* v_{i,n} + Cx_n - z\big)\big]$
	$y_n = x_{n+1} + \theta_n(x_{n+1} - x_n)$
	$v_{i,n+1} = J_{\sigma_{i,n}B_i^{-1}}[v_{i,n} + \sigma_{i,n}(L_iy_n - r_i)],\ i = 1,\dots,m$
	$\tau_{n+1} = \theta_n\tau_n,\ \theta_{n+1} = 1/\sqrt{1 + \tau_{n+1}(2\gamma - \eta\tau_{n+1})/\lambda},$
	$\sigma_{i,n+1} = \sigma_{i,n}/\theta_{n+1},\ i = 1,\dots,m.$

Remark 6 Notice that in contrast to the algorithm in Theorem 2, we allow here variable step sizes τ_n and $\sigma_{i,n}$, $1 = 1,\dots,m$, which are updated in each iteration. Moreover, for every $n \geq 0$ the iterate y_n is defined by using the sequence θ_n. Dynamically adjusted step sizes have been first proposed in [29] and then used in [11] in order to accelerate the convergence of iterative methods when solving convex optimization problems.

Remark 7 Notice that the assumption $\tau_0 \sum_{i=1}^m \sigma_{i,0}\|L_i\|^2 \le \sqrt{1+\tau_0(2\gamma-\eta\tau_0)/\lambda}$ in Algorithm 5 is equivalent to $\tau_1 \sum_{i=1}^m \sigma_{i,0}\|L_i\|^2 \le 1$, being fulfilled if $\tau_0 > 0$ is chosen such that

$$\tau_0 \le \frac{\gamma/\lambda + \sqrt{\gamma^2/\lambda^2 + (\sum_{i=1}^m \sigma_{i,0}\|L_i\|^2)^2 + \eta/\lambda}}{(\sum_{i=1}^m \sigma_{i,0}\|L_i\|^2)^2 + \eta/\lambda}.$$

Theorem 8 *Suppose that $A+C$ is γ-strongly monotone for $\gamma > 0$ and let $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ be a primal-dual solution to Problem 3. Then the sequences generated by Algorithm 5 fulfill for any $n \ge 0$*

$$\frac{\lambda\|x_{n+1}-\overline{x}\|^2}{\tau_{n+1}^2} + \left(1 - \tau_1 \sum_{i=1}^m \sigma_{i,0}\|L_i\|^2\right) \sum_{i=1}^m \frac{\|v_{i,n}-\overline{v}_i\|^2}{\tau_1\sigma_{i,0}} \le$$
$$\frac{\lambda\|x_1-\overline{x}\|^2}{\tau_1^2} + \sum_{i=1}^m \frac{\|v_{i,0}-\overline{v}_i\|^2}{\tau_1\sigma_{i,0}} + \frac{\|x_1-x_0\|^2}{\tau_0^2} + \frac{2}{\tau_0}\sum_{i=1}^m \langle L_i(x_1-x_0), v_{i,0}-\overline{v}_i\rangle.$$

Moreover, $\lim_{n\to+\infty} n\tau_n = \frac{\lambda}{\gamma}$, hence one obtains for $(x_n)_{n\ge0}$ an order of convergence of $\mathcal{O}(\frac{1}{n})$.

Proof. The idea of the proof relies on showing that the following Fejér-type inequality is true for any $n \ge 0$

$$\begin{aligned}
&\frac{\lambda}{\tau_{n+2}^2}\|x_{n+2}-\overline{x}\|^2 + \sum_{i=1}^m \frac{\|v_{i,n+1}-\overline{v}_i\|^2}{\tau_1\sigma_{i,0}} + \frac{\|x_{n+2}-x_{n+1}\|^2}{\tau_{n+1}^2} - \\
&\frac{2}{\tau_{n+1}}\sum_{i=1}^m \langle L_i(x_{n+2}-x_{n+1}), -v_{i,n+1}+\overline{v}_i\rangle \le \\
&\frac{\lambda}{\tau_{n+1}^2}\|x_{n+1}-\overline{x}\|^2 + \sum_{i=1}^m \frac{\|v_{i,n}-\overline{v}_i\|^2}{\tau_1\sigma_{i,0}} + \frac{\|x_{n+1}-x_n\|^2}{\tau_n^2} - \\
&\frac{2}{\tau_n}\sum_{i=1}^m \langle L_i(x_{n+1}-x_n), -v_{i,n}+\overline{v}_i\rangle.
\end{aligned} \tag{8}$$

To this end we use first that in the light of the definition of the resolvents it holds for any $n \ge 0$

$$\frac{\lambda}{\tau_{n+1}}(x_{n+1}-x_{n+2}) - \left(\sum_{i=1}^m L_i^* v_{i,n+1} + Cx_{n+1} - z\right) + Cx_{n+2} \in (A+C)x_{n+2}. \tag{9}$$

Since $A+C$ is γ-strongly monotone, (7) and (9) yield for any $n \ge 0$

$$\begin{aligned}
&\gamma\|x_{n+2}-\overline{x}\|^2 \le \left\langle x_{n+2}-\overline{x}, \frac{\lambda}{\tau_{n+1}}(x_{n+1}-x_{n+2})\right\rangle + \\
&\left\langle x_{n+2}-\overline{x}, -\left(\sum_{i=1}^m L_i^* v_{i,n+1} + Cx_{n+1} - z\right) + Cx_{n+2} - \left(z - \sum_{i=1}^m L_i^*\overline{v}_i\right)\right\rangle = \\
&\frac{\lambda}{\tau_{n+1}}\langle x_{n+2}-\overline{x}, x_{n+1}-x_{n+2}\rangle + \langle x_{n+2}-\overline{x}, Cx_{n+2}-Cx_{n+1}\rangle + \\
&\sum_{i=1}^m \langle L_i(x_{n+2}-\overline{x}), \overline{v}_i - v_{i,n+1}\rangle.
\end{aligned} \tag{10}$$

Further, we have

$$\langle x_{n+2} - \overline{x}, x_{n+1} - x_{n+2}\rangle = \frac{\|x_{n+1} - \overline{x}\|^2}{2} - \frac{\|x_{n+2} - \overline{x}\|^2}{2} - \frac{\|x_{n+1} - x_{n+2}\|^2}{2} \tag{11}$$

and, since C is η-Lipschitzian,

$$\begin{aligned}\langle x_{n+2} - \overline{x}, Cx_{n+2} - Cx_{n+1}\rangle &\leq \|x_{n+2} - \overline{x}\| \cdot \|Cx_{n+2} - Cx_{n+1}\| \\ &\leq \frac{\eta\tau_{n+1}}{2}\|x_{n+2} - \overline{x}\|^2 + \frac{\eta}{2\tau_{n+1}}\|x_{n+2} - x_{n+1}\|^2. \end{aligned}\tag{12}$$

Hence, it follows from (10)–(12) that for any $n \geq 0$, it holds

$$\left(\frac{\lambda}{\tau_{n+1}} + 2\gamma - \eta\tau_{n+1}\right)\|x_{n+2} - \overline{x}\|^2 \leq$$
$$\frac{\lambda}{\tau_{n+1}}\|x_{n+1} - \overline{x}\|^2 - \frac{\lambda - \eta}{\tau_{n+1}}\|x_{n+2} - x_{n+1}\|^2 + 2\sum_{i=1}^{m}\langle L_i(x_{n+2} - \overline{x}), \overline{v}_i - v_{i,n+1}\rangle.$$

Taking into account that $\lambda \geq \eta + 1$, we obtain for any $n \geq 0$ that

$$\left(\frac{\lambda}{\tau_{n+1}} + 2\gamma - \eta\tau_{n+1}\right)\|x_{n+2} - \overline{x}\|^2 \leq$$
$$\frac{\lambda}{\tau_{n+1}}\|x_{n+1} - \overline{x}\|^2 - \frac{1}{\tau_{n+1}}\|x_{n+2} - x_{n+1}\|^2 + 2\sum_{i=1}^{m}\langle L_i(x_{n+2} - \overline{x}), \overline{v}_i - v_{i,n+1}\rangle. \tag{13}$$

On the other hand, for every $i = 1,..., m$ and any $n \geq 0$, from

$$\frac{1}{\sigma_{i,n}}(v_{i,n} - v_{i,n+1}) + L_i y_n - r_i \in B_i^{-1} v_{i,n+1}, \tag{14}$$

the monotonicity of B_i^{-1} and (7), we obtain

$$\begin{aligned} 0 &\leq \left\langle \frac{1}{\sigma_{i,n}}(v_{i,n} - v_{i,n+1}) + L_i y_n - r_i - (L_i\overline{x} - r_i), v_{i,n+1} - \overline{v}_i \right\rangle \\ &= \frac{1}{\sigma_{i,n}}\langle v_{i,n} - v_{i,n+1}, v_{i,n+1} - \overline{v}_i\rangle + \langle L_i(y_n - \overline{x}), v_{i,n+1} - \overline{v}_i\rangle \\ &= \frac{1}{2\sigma_{i,n}}\|v_{i,n} - \overline{v}_i\|^2 - \frac{1}{2\sigma_{i,n}}\|v_{i,n} - v_{i,n+1}\|^2 - \frac{1}{2\sigma_{i,n}}\|v_{i,n+1} - \overline{v}_i\|^2 \\ &+ \langle L_i(y_n - \overline{x}), v_{i,n+1} - \overline{v}_i\rangle, \end{aligned}$$

hence

$$\frac{\|v_{i,n+1} - \overline{v}_i\|^2}{\sigma_{i,n}} \leq \frac{\|v_{i,n} - \overline{v}_i\|^2}{\sigma_{i,n}} - \frac{\|v_{i,n} - v_{i,n+1}\|^2}{\sigma_{i,n}} + 2\langle L_i(y_n - \overline{x}), v_{i,n+1} - \overline{v}_i\rangle. \tag{15}$$

Summing up the inequalities in (13) and (15) we obtain for any $n \geq 0$

$$\left(\frac{\lambda}{\tau_{n+1}} + 2\gamma - \eta\tau_{n+1}\right)\|x_{n+2} - \overline{x}\|^2 + \sum_{i=1}^{m}\frac{\|v_{i,n+1} - \overline{v}_i\|^2}{\sigma_{i,n}} \leq$$
$$\frac{\lambda}{\tau_{n+1}}\|x_{n+1} - \overline{x}\|^2 + \sum_{i=1}^{m}\frac{\|v_{i,n} - \overline{v}_i\|^2}{\sigma_{i,n}} - \frac{\|x_{n+2} - x_{n+1}\|^2}{\tau_{n+1}} - \sum_{i=1}^{m}\frac{\|v_{i,n} - v_{i,n+1}\|^2}{\sigma_{i,n}} \tag{16}$$
$$+2\sum_{i=1}^{m}\langle L_i(x_{n+2} - y_n), -v_{i,n+1} + \overline{v}_i\rangle.$$

Further, since $y_n = x_{n+1} + \theta_n(x_{n+1} - x_n)$, for every $i = 1,...,m$ and any $n \geq 0$, it holds

$$\langle L_i(x_{n+2} - y_n), -v_{i,n+1} + \overline{v}_i\rangle = \langle L_i\big(x_{n+2} - x_{n+1} - \theta_n(x_{n+1} - x_n)\big), -v_{i,n+1} + \overline{v}_i\rangle =$$
$$\langle L_i(x_{n+2} - x_{n+1}), -v_{i,n+1} + \overline{v}_i\rangle - \theta_n\langle L_i(x_{n+1} - x_n), -v_{i,n} + \overline{v}_i\rangle +$$
$$\theta_n\langle L_i(x_{n+1} - x_n), -v_{i,n} + v_{i,n+1}\rangle \leq$$
$$\langle L_i(x_{n+2} - x_{n+1}), -v_{i,n+1} + \overline{v}_i\rangle - \theta_n\langle L_i(x_{n+1} - x_n), -v_{i,n} + \overline{v}_i\rangle +$$
$$\frac{\theta_n^2\|L_i\|^2\sigma_{i,n}}{2}\|x_{n+1} - x_n\|^2 + \frac{\|v_{i,n} - v_{i,n+1}\|^2}{2\sigma_{i,n}}.$$

By combining the last inequality with (16), we obtain for any $n \geq 0$

$$\begin{aligned}
&\left(\frac{\lambda}{\tau_{n+1}} + 2\gamma - \eta\tau_{n+1}\right)\|x_{n+2} - \overline{x}\|^2 + \sum_{i=1}^{m}\frac{\|v_{i,n+1} - \overline{v}_i\|^2}{\sigma_{i,n}} + \frac{\|x_{n+2} - x_{n+1}\|^2}{\tau_{n+1}} - \\
&2\sum_{i=1}^{m}\langle L_i(x_{n+2} - x_{n+1}), -v_{i,n+1} + \overline{v}_i\rangle \leq \\
&\frac{\lambda}{\tau_{n+1}}\|x_{n+1} - \overline{x}\|^2 + \sum_{i=1}^{m}\frac{\|v_{i,n} - \overline{v}_i\|^2}{\sigma_{i,n}} + \left(\sum_{i=1}^{m}\|L_i\|^2\sigma_{i,n}\right)\theta_n^2\|x_{n+1} - x_n\|^2 - \\
&2\sum_{i=1}^{m}\theta_n\langle L_i(x_{n+1} - x_n), -v_{i,n} + \overline{v}_i\rangle.
\end{aligned} \tag{17}$$

After dividing (17) by τ_{n+1} and noticing that for any $n \geq 0$

$$\frac{\lambda}{\tau_{n+1}^2} + \frac{2\gamma}{\tau_{n+1}} - \eta = \frac{\lambda}{\tau_{n+2}^2},$$

$$\tau_{n+1}\sigma_{i,n} = \tau_n\sigma_{i,n-1} = ... = \tau_1\sigma_{i,0}$$

and

$$\frac{\left(\sum_{i=1}^m \|L_i\|^2\sigma_{i,n}\right)\theta_n^2}{\tau_{n+1}} = \frac{\tau_{n+1}\sum_{i=1}^m \|L_i\|^2\sigma_{i,n}}{\tau_n^2} = \frac{\tau_1\sum_{i=1}^m \|L_i\|^2\sigma_{i,0}}{\tau_n^2} \leq \frac{1}{\tau_n^2},$$

it follows that the Fejér-type inequality (8) is true.

Let $N \in \mathbb{N}, N \geq 2$. Summing up the inequality in (8) from $n = 0$ to $N - 1$, it yields

$$\begin{aligned}
&\frac{\lambda}{\tau_{N+1}^2}\|x_{N+1} - \overline{x}\|^2 + \sum_{i=1}^{m}\frac{\|v_{i,N} - \overline{v}_i\|^2}{\tau_1\sigma_{i,0}} + \frac{\|x_{N+1} - x_N\|^2}{\tau_N^2} \leq \\
&\frac{\lambda}{\tau_1^2}\|x_1 - \overline{x}\|^2 + \sum_{i=1}^{m}\frac{\|v_{i,0} - \overline{v}_i\|^2}{\tau_1\sigma_{i,0}} + \frac{\|x_1 - x_0\|^2}{\tau_0^2} + \\
&2\sum_{i=1}^{m}\left(\frac{1}{\tau_N}\langle L_i(x_{N+1} - x_N), -v_{i,N} + \overline{v}_i\rangle - \frac{1}{\tau_0}\langle L_i(x_1 - x_0), -v_{i,0} + \overline{v}_i\rangle\right).
\end{aligned} \tag{18}$$

Further, for every $i = 1,...,m$ we use the inequality

$$\begin{aligned}
&\frac{2}{\tau_N}\langle L_i(x_{N+1} - x_N), -v_{i,N} + \overline{v}_i\rangle \leq \\
&\frac{\sigma_{i,0}\|L_i\|^2}{\tau_N^2(\sum_{i=1}^m \sigma_{i,0}\|L_i\|^2)}\|x_{N+1} - x_N\|^2 + \frac{\sum_{i=1}^m \sigma_{i,0}\|L_i\|^2}{\sigma_{i,0}}\|v_{i,N} - \overline{v}_i\|^2
\end{aligned}$$

and obtain from (18) that

$$\frac{\lambda\|x_{N+1}-\overline{x}\|^2}{\tau_{N+1}^2}+\sum_{i=1}^m\frac{\|v_{i,N}-\overline{v}_i\|^2}{\tau_1\sigma_{i,0}}\leq\frac{\lambda\|x_1-\overline{x}\|^2}{\tau_1^2}+\sum_{i=1}^m\frac{\|v_{i,0}-\overline{v}_i\|^2}{\tau_1\sigma_{i,0}}+\frac{\|x_1-x_0\|^2}{\tau_0^2}$$

$$+\frac{2}{\tau_0}\sum_{i=1}^m\langle L_i(x_1-x_0),v_{i,0}-\overline{v}_i\rangle+\sum_{i=1}^m\frac{\sum_{j=1}^m\sigma_{j,0}\|L_j\|^2}{\sigma_{i,0}}\|v_{i,N}-\overline{v}_i\|^2,$$

which rapidly yields the inequality in the statement of the theorem.

We close the proof by showing that $\lim\limits_{n\to+\infty}n\tau_n=\lambda/\gamma$. Notice that for any $n\geq 0$,

$$\tau_{n+1}=\frac{\tau_n}{\sqrt{1+\frac{\tau_n}{\lambda}(2\gamma-\eta\tau_n)}}. \tag{19}$$

Since $0<\tau_0<2\gamma/\eta$, it follows by induction that $0<\tau_{n+1}<\tau_n<\tau_0<2\gamma/\eta$ for any $n\geq 1$, hence the sequence $(\tau_n)_{n\geq 0}$ converges. In the light of (19) one easily obtains that $\lim\limits_{n\to+\infty}\tau_n=0$ and, further, that $\lim\limits_{n\to+\infty}\frac{\tau_n}{\tau_{n+1}}=1$. As $(\frac{1}{\tau_n})_{n\geq 0}$ is a strictly increasing and unbounded sequence, by applying the Stolz-Cesàro Theorem, it yields

$$\begin{aligned}\lim_{n\to+\infty}n\tau_n&=\lim_{n\to+\infty}\frac{n}{\frac{1}{\tau_n}}=\lim_{n\to+\infty}\frac{n+1-n}{\frac{1}{\tau_{n+1}}-\frac{1}{\tau_n}}=\lim_{n\to+\infty}\frac{\tau_n\tau_{n+1}}{\tau_n-\tau_{n+1}}\\&=\lim_{n\to+\infty}\frac{\tau_n\tau_{n+1}(\tau_n+\tau_{n+1})}{\tau_n^2-\tau_{n+1}^2}=\lim_{n\to+\infty}\frac{\tau_n\tau_{n+1}(\tau_n+\tau_{n+1})}{\tau_{n+1}^2\frac{\tau_n}{\lambda}(2\gamma-\eta\tau_n)}\\&=\lim_{n\to+\infty}\frac{\tau_n+\tau_{n+1}}{\tau_{n+1}(\frac{2\gamma}{\lambda}-\frac{\eta}{\lambda}\tau_n)}=\lim_{n\to+\infty}\frac{\frac{\tau_n}{\tau_{n+1}}+1}{\frac{2\gamma}{\lambda}-\frac{\eta}{\lambda}\tau_n}=\frac{\lambda}{\gamma}.\end{aligned}$$

■

Remark 9 Let us mention that, if $A+C$ is γ-strongly monotone with $\gamma>0$, then the operator $A+\sum_{i=1}^m L_i^*(B_i(L_i\cdot-r_i))+C$ is strongly monotone, as well, thus the monotone inclusion problem (5) has at most one solution. Hence, if $(\overline{x},\overline{v}_1,...,\overline{v}_m)$ is a primal-dual solution to Problem 3, then $\overline{x}$ is the unique solution to (5). Notice that the problem (6) may not have an unique solution.

2.2 The case $A+C$ and $B_i^{-1}+D_i^{-1}$, $i=1,...,m$, are strongly monotone

In this subsection we propose a modified version of the algorithm in Theorem 2 which guarantees when $A+C$ and $B_i^{-1}+D_i^{-1}$, $i=1,...,m$, are strongly monotone orders of convergence of $\mathcal{O}(\omega^n)$, for $\omega\in(0,1)$, for the sequences $(x_n)_{n\geq 0}$ and $(v_{i,n})_{n\geq 0}$, $i=1,...,m$. The algorithm aims to solve the primal-dual pair of monotone inclusions stated in Problem 1 under relaxed assumptions for the operators C and D_i^{-1}, $i=1,...,m$. A same comment as in Remark 15 can be made also in this context.

Problem 10 Let $\mathcal{H}$ be a real Hilbert space, $z\in\mathcal{H}$, $A:\mathcal{H}\rightrightarrows\mathcal{H}$ a maximally monotone operator and $C:\mathcal{H}\to\mathcal{H}$ a monotone and η-Lipschitzian operator for $\eta>0$. Let m be a strictly positive integer and, for any $i\in\{1,...,m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i\in\mathcal{G}_i$, let $B_i:\mathcal{G}_i\rightrightarrows\mathcal{G}_i$ be a maximally monotone operator, let $D_i:\mathcal{G}_i\rightrightarrows\mathcal{G}_i$ be a monotone operator such that D_i^{-1} is ν_i-Lipschitzian for $\nu_i > 0$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^*\big((B_i \Box D_i)(L_i\overline{x} - r_i)\big) + C\overline{x}, \tag{20}$$

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,\ldots, \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \begin{cases} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in (B_i \Box D_i)(L_i x - r_i),\ i = 1,\ldots, m. \end{cases} \tag{21}$$

Under the assumption that $A + C$ is γ-strongly monotone for $\gamma > 0$ and $B_i^{-1} + D_i^{-1}$ is δ_i-strongly monotone with $\delta_i > 0$, $i = 1,\ldots.m$, we propose the following modification of the iterative scheme in Theorem 2.

Algorithm 11

`Initialization:` `Choose` $\mu > 0$ `such that`
$\mu \leq \min\left\{\gamma^2/\eta^2, \delta_1^2/\nu_1^2,\ldots, \delta_m^2/\nu_m^2, \sqrt{\gamma/\left(\sum_{i=1}^{m} \|L_i\|^2/\delta_i\right)}\right\}$,
$\tau = \mu/(2\gamma)$, $\sigma_i = \mu/(2\delta_i)$, $i = 1,..,m$,
$\theta \in [2/(2 + \mu), 1]$ `and` $(x_0, v_{1,0},\ldots, v_{m,0}) \in \mathcal{H}\times\ \mathcal{G}_1\ \times\ldots\times\ \mathcal{G}_m$.

`For` $n \geq 0$ `set:` $x_{n+1} = J_{\tau A}\big[x_n - \tau\big(\sum_{i=1}^{m} L_i^* v_{i,n} + Cx_n - z\big)\big]$
$y_n = x_{n+1} + \theta(x_{n+1} - x_n)$
$v_{i,n+1} = J_{\sigma_i B_i^{-1}}[v_{i,n} + \sigma_i(L_i y_n - D_i^{-1} v_{i,n} - r_i)]$, $i = 1,\ldots, m$.

Remark 12 Different to Algorithm 5, the step sizes are now constant in each iteration, as it is also the case in Theorem 2. The major difference to the iterative scheme in Theorem 2 is given by the role played by the constant μ, not only in the definition of the step sizes, but also in the way the sequence $(y_n)_{n\geq 0}$ is constructed (through the choice of θ). Notice that the situation when $\theta = 1$ provides the same definition of the latter as in the algorithm stated in Theorem 2.

Theorem 13 *Suppose that $A + C$ is γ-strongly monotone for $\gamma > 0$, $B_i^{-1} + D_i^{-1}$ is δ_i-strongly monotone for $\delta_i > 0$, $i = 1,\ldots, m$, and let $(\overline{x}, \overline{v}_1,\ldots, \overline{v}_m)$ be a primal-dual solution to Problem 10. Then the sequences generated by Algorithm 11 fulfill for any $n \geq 0$*

$$\gamma\|x_{n+1} - \overline{x}\|^2 + (1 - \omega)\sum_{i=1}^{m} \delta_i\|v_{i,n} - \overline{v}_i\|^2 \leq$$
$$\omega^n \left(\gamma\|x_1 - \overline{x}\|^2 + \sum_{i=1}^{m} \delta_i\|v_{i,0} - \overline{v}_i\|^2 + \frac{\gamma}{2}\omega\|x_1 - x_0\|^2 + \mu\omega\sum_{i=1}^{m}\langle L_i(x_1 - x_0), v_{i,0} - \overline{v}_i\rangle\right),$$

where $0 < \omega = \frac{2(1+\theta)}{4+\mu} < 1$.

Proof. For any $n \geq 0$ we have

$$\frac{1}{\tau}(x_{n+1} - x_{n+2}) - \left(\sum_{i=1}^{m} L_i^* v_{i,n+1} + Cx_{n+1} - z\right) + Cx_{n+2} \in (A + C)x_{n+2}, \tag{22}$$

thus, since $A + C$ is γ-strongly monotone, (21) yields

$$\gamma\|x_{n+2} - \overline{x}\|^2 \leq \left\langle x_{n+2} - \overline{x}, \frac{1}{\tau}(x_{n+1} - x_{n+2}) \right\rangle +$$
$$\left\langle x_{n+2} - \overline{x}, -\left(\sum_{i=1}^{m} L_i^* v_{i,n+1} + Cx_{n+1} - z\right) + Cx_{n+2} - \left(z - \sum_{i=1}^{m} L_i^* \overline{v}_i\right)\right\rangle = \quad (23)$$
$$\frac{1}{\tau}\langle x_{n+2} - \overline{x}, x_{n+1} - x_{n+2}\rangle + \langle x_{n+2} - \overline{x}, Cx_{n+2} - Cx_{n+1}\rangle +$$
$$\sum_{i=1}^{m} \langle L_i(x_{n+2} - \overline{x}), \overline{v}_i - v_{i,n+1}\rangle .$$

Further, by using (11) and

$$\langle x_{n+2} - \overline{x}, Cx_{n+2} - Cx_{n+1}\rangle \leq \frac{\gamma}{2}\|x_{n+2} - \overline{x}\|^2 + \frac{\eta^2}{2\gamma}\|x_{n+2} - x_{n+1}\|^2,$$

we get from (23) that for any $n \geq 0$

$$\left(\frac{1}{2\tau} + \frac{\gamma}{2}\right)\|x_{n+2} - \overline{x}\|^2 \leq$$
$$\frac{1}{2\tau}\|x_{n+1} - \overline{x}\|^2 - \left(\frac{1}{2\tau} - \frac{\eta^2}{2\gamma}\right)\|x_{n+2} - x_{n+1}\|^2 + \sum_{i=1}^{m}\langle L_i(x_{n+2} - \overline{x}), \overline{v}_i - v_{i,n+1}\rangle.$$

After multiplying this inequality with μ and taking into account that

$$\frac{\mu}{2\tau} = \gamma, \mu\left(\frac{1}{2\tau} + \frac{\gamma}{2}\right) = \gamma\left(1 + \frac{\mu}{2}\right) \text{ and } \mu\left(\frac{1}{2\tau} - \frac{\eta^2}{2\gamma}\right) = \gamma - \frac{\eta^2}{2\gamma}\mu \geq \frac{\gamma}{2},$$

we obtain for any $n \geq 0$

$$\gamma\left(1 + \frac{\mu}{2}\right)\|x_{n+2} - \overline{x}\|^2 \leq \quad (24)$$
$$\gamma\|x_{n+1} - \overline{x}\|^2 - \frac{\gamma}{2}\|x_{n+2} - x_{n+1}\|^2 + \mu\sum_{i=1}^{m}\langle L_i(x_{n+2} - \overline{x}), \overline{v}_i - v_{i,n+1}\rangle.$$

On the other hand, for every $i = 1,...,m$ and any $n \geq 0$, from

$$\frac{1}{\sigma_i}(v_{i,n} - v_{i,n+1}) + L_i y_n - D_i^{-1}v_{i,n} - r_i + D_i^{-1}v_{i,n+1} \in (B_i^{-1} + D_i^{-1})v_{i,n+1}, \quad (25)$$

the δ_i-strong monotonicity of $B_i^{-1} + D_i^{-1}$ and (21), we obtain

$$\begin{aligned}\delta_i\|v_{i,n+1} - \overline{v}_i\|^2 \quad &\leq \quad \left\langle \frac{1}{\sigma_i}(v_{i,n} - v_{i,n+1}), v_{i,n+1} - \overline{v}_i \right\rangle \\ &+ \quad \left\langle L_i y_n - r_i - D_i^{-1}v_{i,n} + D_i^{-1}v_{i,n+1} - (L_i\overline{x} - r_i), v_{i,n+1} - \overline{v}_i \right\rangle . (26)\end{aligned}$$

Further, for every $i = 1,...,m$ and any $n \geq 0$, we have

$$\frac{1}{\sigma_i}\langle v_{i,n} - v_{i,n+1}, v_{i,n+1} - \overline{v}_i\rangle = \frac{1}{2\sigma_i}\|v_{i,n} - \overline{v}_i\|^2 - \frac{1}{2\sigma_i}\|v_{i,n} - v_{i,n+1}\|^2 - \frac{1}{2\sigma_i}\|v_{i,n+1} - \overline{v}_i\|^2$$

and, since D_i^{-1} is a ν_i-Lipschitzian operator,

$$\langle D_i^{-1}v_{i,n+1} - D_i^{-1}v_{i,n}, v_{i,n+1} - \overline{v}_i\rangle \leq \frac{\delta_i}{2}\|v_{i,n+1} - \overline{v}_i\|^2 + \frac{\nu_i^2}{2\delta_i}\|v_{i,n+1} - v_{i,n}\|^2. \tag{27}$$

Consequently, from (26) and(27) we obtain for every $i = 1,..., m$ and any $n \geq 0$:

$$\left(\frac{1}{2\sigma_i} + \frac{\delta_i}{2}\right)\|v_{i,n+1} - \overline{v}_i\|^2 \leq$$
$$\frac{1}{2\sigma_i}\|v_{i,n} - \overline{v}_i\|^2 - \left(\frac{1}{2\sigma_i} - \frac{\nu_i^2}{2\delta_i}\right)\|v_{i,n+1} - v_{i,n}\|^2 + \langle L_i(\overline{x} - y_n), \overline{v}_i - v_{i,n+1}\rangle,$$

which, after multiplying it by μ (here is the initial choice of μ determinant), yields

$$\delta_i\left(1 + \frac{\mu}{2}\right)\|v_{i,n+1}-\overline{v}_i\|^2 \leq \delta_i\|v_{i,n}-\overline{v}_i\|^2-\frac{\delta_i}{2}\|v_{i,n+1}-v_{i,n}\|^2+\mu\langle L_i(\overline{x}-y_n), \overline{v}_i-v_{i,n+1}\rangle. \tag{28}$$

We denote

$$a_n := \gamma\|x_{n+1} - \overline{x}\|^2 + \sum_{i=1}^m \delta_i\|v_{i,n} - \overline{v}_i\|^2 \ \forall n \geq 0.$$

Summing up the inequalities in (24) and (28), we obtain for any $n \geq 0$

$$\left(1 + \frac{\mu}{2}\right)a_{n+1} \leq a_n \tag{29}$$
$$-\frac{\gamma}{2}\|x_{n+2} - x_{n+1}\|^2 - \sum_{i=1}^m \frac{\delta_i}{2}\|v_{i,n} - v_{i,n+1}\|^2 + \mu\sum_{i=1}^m\langle L_i(x_{n+2} - y_n), \overline{v}_i - v_{i,n+1}\rangle.$$

Further, since $y_n = x_{n+1} + \theta(x_{n+1} - x_n)$ and $\omega \leq \theta$, for every $i = 1,..., m$ and any $n \geq 0$, it holds

$$\langle L_i(x_{n+2} - y_n), \overline{v}_i - v_{i,n+1}\rangle = \langle L_i\left(x_{n+2} - x_{n+1} - \theta(x_{n+1} - x_n)\right), \overline{v}_i - v_{i,n+1}\rangle =$$
$$\langle L_i(x_{n+2} - x_{n+1}), \overline{v}_i - v_{i,n+1}\rangle - \omega\langle L_i(x_{n+1} - x_n), \overline{v}_i - v_{i,n}\rangle +$$
$$\omega\langle L_i(x_{n+1} - x_n), v_{i,n+1} - v_{i,n}\rangle + (\theta - \omega)\langle L_i(x_{n+1} - x_n), v_{i,n+1} - \overline{v}_i\rangle \leq$$
$$\langle L_i(x_{n+2} - x_{n+1}), \overline{v}_i - v_{i,n+1}\rangle - \omega\langle L_i(x_{n+1} - x_n), \overline{v}_i - v_{i,n}\rangle +$$
$$\omega\|L_i\|\left(\mu\omega\|L_i\|\frac{\|x_{n+1} - x_n\|^2}{2\delta_i} + \delta_i\frac{\|v_{i,n+1} - v_{i,n}\|^2}{2\mu\omega\|L_i\|}\right) +$$
$$(\theta - \omega)\|L_i\|\left(\mu\omega\|L_i\|\frac{\|x_{n+1} - x_n\|^2}{2\delta_i} + \delta_i\frac{\|v_{i,n+1} - \overline{v}_i\|^2}{2\mu\omega\|L_i\|}\right) =$$
$$\langle L_i(x_{n+2} - x_{n+1}), \overline{v}_i - v_{i,n+1}\rangle - \omega\langle L_i(x_{n+1} - x_n), \overline{v}_i - v_{i,n}\rangle +$$
$$\theta\mu\omega\|L_i\|^2\frac{\|x_{n+1} - x_n\|^2}{2\delta_i} + \delta_i\frac{\|v_{i,n+1} - v_{i,n}\|^2}{2\mu} + (\theta - \omega)\delta_i\frac{\|v_{i,n+1} - \overline{v}_i\|^2}{2\mu\omega}.$$

Taking into consideration that

$$\frac{\mu^2\theta\omega}{2}\sum_{i=1}^m\frac{\|L_i\|^2}{\delta_i} \leq \frac{\gamma\theta}{2}\omega \leq \frac{\gamma}{2}\omega \text{ and } 1 + \frac{\mu}{2} = \frac{1}{\omega} + \frac{\theta - \omega}{\omega},$$

from (29), we obtain for any $n \geq 0$

$$\begin{aligned}
&\frac{1}{\omega}a_{n+1} + \frac{\gamma}{2}\|x_{n+2} - x_{n+1}\|^2 \leq \\
&a_n + \frac{\gamma}{2}\omega\|x_{n+1} - x_n\|^2 - \frac{\theta - \omega}{\omega}\left(a_{n+1} - \sum_{i=1}^{m}\frac{\delta_i}{2}\|v_{i,n+1} - \overline{v}_i\|^2\right) + \\
&\mu\sum_{i=1}^{m}\big(\langle L_i(x_{n+2} - x_{n+1}), \overline{v}_i - v_{i,n+1}\rangle - \omega\langle L_i(x_{n+1} - x_n), \overline{v}_i - v_{i,n}\rangle\big).
\end{aligned}$$

As $\omega \leq \theta$ and $a_{n+1} - \sum_{i=1}^{m}\frac{\delta_i}{2}\|v_{i,n+1} - \overline{v}_i\|^2 \geq 0$, we further get after multiplying the last inequality with ω^{-n} the following Fejér-type inequality that holds for any $n \geq 0$

$$\begin{aligned}
&\omega^{-(n+1)}a_{n+1} + \frac{\gamma}{2}\omega^{-n}\|x_{n+2} - x_{n+1}\|^2 + \mu\omega^{-n}\sum_{i=1}^{m}\langle L_i(x_{n+2} - x_{n+1}), v_{i,n+1} - \overline{v}_i\rangle \leq \\
&\omega^{-n}a_n + \frac{\gamma}{2}\omega^{-(n-1)}\|x_{n+1} - x_n\|^2 + \mu\omega^{-(n-1)}\sum_{i=1}^{m}\langle L_i(x_{n+1} - x_n), v_{i,n} - \overline{v}_i\rangle. \qquad (30)
\end{aligned}$$

Let $N \in \mathbb{N}, N \geq 2$. Summing up the inequality in (30) from $n = 0$ to $N - 1$, it yields

$$\begin{aligned}
&\omega^{-N}a_N + \frac{\gamma}{2}\omega^{-N+1}\|x_N - x_{N+1}\|^2 + \mu\omega^{-N+1}\sum_{i=1}^{m}\langle L_i(x_{N+1} - x_N), v_{i,N} - \overline{v}_i\rangle \leq \\
&a_0 + \frac{\gamma}{2}\omega\|x_1 - x_0\|^2 + \mu\omega\sum_{i=1}^{m}\langle L_i(x_1 - x_0), v_{i,0} - \overline{v}_i\rangle.
\end{aligned}$$

Using that

$$\langle L_i(x_{N+1} - x_N), v_{i,N} - \overline{v}_i\rangle \geq -\frac{\mu\|L_i\|^2}{4\delta_i}\|x_{N+1} - x_N\|^2 - \frac{\delta_i}{\mu}\|v_{i,N} - \overline{v}_i\|^2,\ i = 1,\ldots,m,$$

this further yields

$$\begin{aligned}
&\omega^{-N}a_N + \omega^{-N+1}\left(\frac{\gamma}{2} - \frac{\mu^2}{4}\sum_{i=1}^{m}\frac{\|L_i\|^2}{\delta_i}\right)\|x_N - x_{N+1}\|^2 - \omega^{-N+1}\sum_{i=1}^{m}\delta_i\|v_{i,N} - \overline{v}_i\|^2 \leq \\
&a_0 + \frac{\gamma}{2}\omega\|x_1 - x_0\|^2 + \mu\omega\sum_{i=1}^{m}\langle L_i(x_1 - x_0), v_{i,0} - \overline{v}_i\rangle. \qquad (31)
\end{aligned}$$

Taking into account the way μ has been chosen, we have

$$\frac{\gamma}{2} - \frac{\mu^2}{4}\sum_{i=1}^{m}\frac{\|L_i\|^2}{\delta_i} \geq \frac{\gamma}{2} - \frac{\gamma}{4} > 0,$$

hence, after multiplying (31) with ω^{-N}, it yields

$$a_N - \omega\sum_{i=1}^{m}\delta_i\|v_{i,N} - \overline{v}_i\|^2 \leq \omega^N\left(a_0 + \frac{\gamma}{2}\omega\|x_1 - x_0\|^2 + \mu\omega\sum_{i=1}^{m}\langle L_i(x_1 - x_0), v_{i,0} - \overline{v}_i\rangle\right).$$

The conclusion follows by taking into account the definition of the sequence $(a_n)_{n\geq 0}$. ■

Remark 14 If $A + C$ is γ-strongly monotone for $\gamma > 0$ and $B_i^{-1} + D_i^{-1}$ is δ_i-strongly monotone for $\delta_i > 0$, $i = 1,...,m$, then there exists at most one primal-dual solution to Problem 10. Hence, if $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ is a primal-dual solution to Problem 10, then $\overline{x}$ is the unique solution to the primal inclusion (20) and $(\overline{v}_1,..., \overline{v}_m)$ is the unique solution to the dual inclusion (21).

Remark 15 The modified versions Algortihm 5 and Algorithm 11 can handle Problem 1 under more general hypotheses than the original method given in [27]. Indeed, convergence was shown under more general hypotheses on the operator C for the first (see also Remark 4) and on the operators $D_i, i = 1, ..., m$ for the latter. More than that, we can provide in both cases a rate of convergence for the sequence of the primal iterates and in case of Algorithm 11 one for the sequence of dual iterates, as well, in particular also strong convergence.

Remark 16 As mentioned in the introduction, in [7] accelerated versions of the algorithm from [14] have been proposed. The algorithms in [7] and the ones proposed in this manuscript are designed to solve the same type of problems and under the same hypotheses concerning the operators involved (compare [7, Theorem 3.3] with Theorem 8 above and [7, Theorem 3.4] with Theorem 13, respectively). The rates of convergence obtained in [7] and in our paper are the same.

On the other hand, our schemes differ from the ones in [7] in some fundamental aspects. Indeed, we propose here accelerated versions of the algorithm given in [27], which relies on a forward-backward scheme, while in [7] the accelerated versions are with respect to a forward-backward-forward scheme. In contrast to the forward-backward-forward algorithm, which requires additional sequences to be computed, the forward-backward scheme needs fewer steps, thus presents from theoretical point of view an important advantage. This applies also for the accelerated versions of these algorithms. The mentioned advantage is underlined also by the numerical results presented in the last section of our paper. Moreover, one can notice that in Algorithm 5 at every iteration when evaluating the operators B_i different step sizes (in form of the parameters $\sigma_{i,n}$) for $i = 1, ..., m$ have been considered, which is not the case with the iterative scheme in [7, Theorem 3.3] where for the evaluation of the same operators the same step size has been used. Individual step sizes possess the advantage that in this way the operators B_i, $i = 1, ..., m$ can be more involved in the algorithm and in the improvement of its convergence properties. A similar remark can be made also for the iterative scheme in [7, Theorem 3.4] and Algorithm 11.

3 Convex optimization problems

The aim of this section is to show that the two algorithms proposed in this paper and investigated from the point of view of their convergence properties can be employed when solving a primal-dual pair of convex optimization problems.

For a function $f : \mathcal{H} \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ is the extended real line, we denote by $\operatorname{dom} f = \{x \in \mathcal{H} : f(x) < +\infty\}$ its *effective domain* and say that f is *proper* if $\operatorname{dom} f \neq \emptyset$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. We denote by $\Gamma(\mathcal{H})$ the family of proper convex and lower semi-continuous extended real-valued functions defined on $\mathcal{H}$. Let $f^* : \mathcal{H} \to \overline{\mathbb{R}}$, $f^*(u) = \sup_{x \in \mathcal{H}}\{\langle u, x\rangle - f(x)\}$ for all $u \in \mathcal{H}$, be the *conjugate function* of f. The *subdifferential* of f at $x \in \mathcal{H}$, with $f(x) \in \mathbb{R}$, is the set $\partial f(x) := \{v \in \mathcal{H} : f(y) \geq f(x) + \langle v, y - x\rangle \ \forall y \in \mathcal{H}\}$.

We take by convention $\partial f(x) := \emptyset$, if $f(x) \in \{\pm\infty\}$. Notice that if $f \in \Gamma(\mathcal{H})$, then ∂f is a maximally monotone operator (cf. [22]) and it holds $(\partial f)^{-1} = \partial f^*$. For $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ two proper functions, we consider their *infimal convolution*, which is the function $f \Box g : \mathcal{H} \to \overline{\mathbb{R}}$, defined by $(f \Box g)(x) = \inf_{y \in \mathcal{H}}\{f(y) + g(x-y)\}$, for all $x \in \mathcal{H}$.

Let $S \subseteq \mathcal{H}$ be a nonempty set. The *indicator function* of S, $\delta_S : \mathcal{H} \to \overline{\mathbb{R}}$, is the function which takes the value 0 on S and $+\infty$ otherwise. The subdifferential of the indicator function is the *normal cone* of S, that is $N_S(x) = \{u \in \mathcal{H} : \langle u, y-x\rangle \leq 0 \ \forall y \in S\}$, if $x \in S$ and $N_S(x) = \emptyset$ for $x \notin S$.

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

$$\inf_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}. \tag{32}$$

Notice that $J_{\gamma \partial f} = (\operatorname{Id}_{\mathcal{H}} + \gamma \partial f)^{-1} = \operatorname{prox}_{\gamma f}$, thus $\operatorname{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H}$ is a single-valued operator fulfilling the extended *Moreau's decomposition formula*

$$\operatorname{prox}_{\gamma f} + \gamma \operatorname{prox}_{(1/\gamma) f^*} \circ \gamma^{-1} \operatorname{Id}_{\mathcal{H}} = \operatorname{Id}_{\mathcal{H}}. \tag{33}$$

Let us also recall that the function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is said to be γ*-strongly convex* for $\gamma > 0$, if $f - \frac{\gamma}{2}\|\cdot\|^2$ is a convex function. Let us mention that this property implies γ-strong monotonicity of ∂f (see [1, Example 22.3]).

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

$$J_{\gamma N_S} = J_{N_S} = J_{\partial \delta_S} = (\operatorname{Id}_{\mathcal{H}} + N_S)^{-1} = \operatorname{prox}_{\delta_S} = P_S, \tag{34}$$

where $P_S : \mathcal{H} \to C$ denotes the *projection operator* on S (see [1, Example 23.3 and Example 23.4]).

In order to investigate the applicability of the algorithm introduced in Subsection 2.1 we consider the following primal-dual pair of convex optimization problems.

Problem 17 Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with a η-Lipschitzian gradient for $\eta > 0$. Let m be a strictly positive integer and, for any $i \in \{1,..., m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, $g_i \in \Gamma(\mathcal{G}_i)$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. Consider the convex optimization problem

$$\inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} g_i(L_i x - r_i) + h(x) - \langle x, z\rangle \right\} \tag{35}$$

and its *Fenchel-type dual* problem

$$\sup_{v_i \in \mathcal{G}_i,\, i=1,..,m} \left\{ -\left(f^* \Box h^*\right)\left(z - \sum_{i=1}^{m} L_i^* v_i\right) - \sum_{i=1}^{m} \left(g_i^*(v_i) + \langle v_i, r_i\rangle\right) \right\}. \tag{36}$$

Considering maximal monotone operators

$$A = \partial f, C = \nabla h \text{ and } B_i = \partial g_i,\ i = 1,..., m,$$

the monotone inclusion problem (5) reads

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in \partial f(\overline{x}) + \sum_{i=1}^{m} L_i^*(\partial g_i(L_i\overline{x} - r_i)) + \nabla h(\overline{x}), \tag{37}$$

while the dual inclusion problem (6) reads

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \begin{cases} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in \partial f(x) + \nabla h(x) \\ \overline{v}_i \in \partial g_i(L_i x - r_i),\ i = 1,..., m. \end{cases} \tag{38}$$

If $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to (37)-(38), namely,

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in \partial g_i(L_i\overline{x} - r_i),\ i = 1,..., m, \tag{39}$$

then $\overline{x}$ is an optimal solution of the problem (35), $(\overline{v}_1,..., \overline{v}_m)$ is an optimal solution of (36) and the optimal objective values of the two problems coincide. Notice that (39) is nothing else than the system of optimality conditions for the primal-dual pair of convex optimization problems (35)-(36).

In case a qualification condition is fulfilled, these optimality conditions are also necessary. For the readers convenience, let us present some qualification conditions which are suitable in this context. One of the weakest qualification conditions of interiority-type reads (see, for instance, [14, Proposition 4.3, Remark 4.4])

$$(r_1,..., r_m) \in \operatorname{sqri}\left(\prod_{i=1}^{m} \operatorname{dom} g_i - \{(L_1x,..., L_mx) : x \in \operatorname{dom} f\}\right). \tag{40}$$

Here, for $\mathcal{H}$ a real Hilbert space and $S \subseteq \mathcal{H}$ a convex set, we denote by

$$\operatorname{sqri} S := \{x \in S : \cup_{\lambda>0}\lambda(S - x) \text{ is a closed linear subspace of } \mathcal{H}\}$$

its *strong quasi-relative interior*. Notice that we always have $\operatorname{int} S \subseteq \operatorname{sqri} S$ (in general this inclusion may be strict). If $\mathcal{H}$ is finite-dimensional, then $\operatorname{sqri} S$ coincides with $\operatorname{ri} S$, the relative interior of S, which is the interior of S with respect to its affine hull. The condition (40) is fulfilled if (i) $\operatorname{dom} g_i = \mathcal{G}_i$, $i = 1,..., m$ or (ii) $\mathcal{H}$ and $\mathcal{G}_i$ are finite-dimensional and there exists $x \in \operatorname{ri}\operatorname{dom} f$ such that $L_i x - r_i \in \operatorname{ri}\operatorname{dom} g_i$, $i = 1,..., m$ (see [14, Proposition 4.3]). Another useful and easily verifiable qualification condition guaranteeing the optimality conditions (39) has the following formulation: there exists $x' \in \operatorname{dom} f \cap \bigcap_{i=1}^{m} L_i^{-1}(r_i + \operatorname{dom} g_i)$ such that g_i is continuous at $L_i x' - r_i$, $i = 1,..., m$ (see [4, Remark 2.5] and [6]). For other qualification conditions we refer the reader to [1, 3–5, 28].

The following two statements are particular instances of Algorithm 5 and Theorem 8, respectively.

Algorithm 18

`Initialization`*:* `Choose` $\tau_0 > 0, \sigma_{i,0} > 0$, $i = 1,..., m$, `such that`
$\tau_0 < 2\gamma/\eta$, $\lambda \geq \eta + 1$, $\tau_0 \sum_{i=1}^{m} \sigma_{i,0}\|L_i\|^2 \leq \sqrt{1 + \tau_0(2\gamma - \eta\tau_0)/\lambda}$
$\theta_0 = 1/\sqrt{1+\tau_0(2\gamma-\eta\tau_0)/\lambda}$ `and` $(x_0, v_{1,0},..., v_{m,0}) \in \mathcal{H}\times\mathcal{G}_1\times...\times\mathcal{G}_m$.

`For` $n \geq 0$ `set`*:*
$x_{n+1} = \operatorname{prox}_{(\tau_n/\lambda)f}\left[x_n - (\tau_n/\lambda)\left(\sum_{i=1}^{m} L_i^* v_{i,n} + \nabla h(x_n) - z\right)\right]$
$y_n = x_{n+1} + \theta_n(x_{n+1} - x_n)$
$v_{i,n+1} = \operatorname{prox}_{\sigma_{i,n} g_i^*}[v_{i,n} + \sigma_{i,n}(L_i y_n - r_i)]$, $i = 1,..., m$
$\tau_{n+1} = \theta_n\tau_n$, $\theta_{n+1} = 1/\sqrt{1 + \tau_{n+1}(2\gamma - \eta\tau_{n+1})/\lambda}$
$\sigma_{i,n+1} = \sigma_{i,n}/\theta_{n+1}$, $i = 1,..., m$.

Theorem 19 *Suppose that $f + h$ is γ-strongly convex for $\gamma > 0$ and the qualification condition* (40) *holds. Then there exists a unique optimal solution $\overline{x}$ to* (35)*, an optimal solution $(\overline{v}_1,..., \overline{v}_m)$ to* (36) *fulfilling the optimality conditions* (39) *and such that the optimal objective values of the problems* (35) *and* (36) *coincide. The sequences generated by Algorithm 18 fulfill for any $n \geq 0$*

$$\frac{\lambda\|x_{n+1} - \overline{x}\|^2}{\tau_{n+1}^2} + \left(1 - \tau_1 \sum_{i=1}^{m} \sigma_{i,0}\|L_i\|^2\right) \sum_{i=1}^{m} \frac{\|v_{i,n} - \overline{v}_i\|^2}{\tau_1\sigma_{i,0}} \leq$$
$$\frac{\lambda\|x_1 - \overline{x}\|^2}{\tau_1^2} + \sum_{i=1}^{m} \frac{\|v_{i,0} - \overline{v}_i\|^2}{\tau_1\sigma_{i,0}} + \frac{\|x_1 - x_0\|^2}{\tau_0^2} + \frac{2}{\tau_0}\sum_{i=1}^{m}\langle L_i(x_1 - x_0), v_{i,0} - \overline{v}_i\rangle.$$

Moreover, $\lim_{n\to+\infty} n\tau_n = \frac{\lambda}{\gamma}$, hence one obtains for $(x_n)_{n\geq0}$ an order of convergence of $\mathcal{O}(\frac{1}{n})$.

Remark 20 The uniqueness of the solution of (35) in the above theorem follows from [1, Corollary 11.16].

Remark 21 In case $h(x) = 0$ for all $x \in \mathcal{H}$, one has to choose in Algorithm 18 as initial points $\tau_0 > 0, \sigma_{i,0} > 0,\ i = 1, .., m$, with $\tau_0 \sum_{i=1}^{m} \sigma_{i,0}\|L_i\|^2 \leq \sqrt{1 + 2\tau_0\gamma/\lambda}$ and $\lambda \geq 1$ and to update the sequence $(\theta_n)_{n\geq0}$ via $\theta_n = 1/\sqrt{1 + 2\tau_n\gamma/\lambda}$ for any $n \geq 0$, in order to obtain a suitable iterative scheme for solving the pair of primal-dual optimization problems (35)-(36) with the same convergence behavior as of Algorithm 18. In this situation, when chosing $\lambda = 1$, $m = 1$, $z = 0$ and $r_i = 0$, one obtains an algorithm which is equivalent to the one presented by Chambolle and Pock in [11, Algorithm 2].

We turn now our attention to the algorithm introduced in Subsection 2.2 and consider to this end the following primal-dual pair of convex optimization problems.

Problem 22 Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with a η-Lipschitzian gradient for $\eta > 0$. Let m be a strictly positive integer and for any $i \in \{1,..., m\}$ let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that l_i is ν_i^{-1}-strongly convex for $\nu_i > 0$ and $L_i : \mathcal{H} \to \mathcal{G}_i$ a nonzero linear continuous operator. Consider the convex optimization problem

$$\inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m}(g_i \Box l_i)(L_i x - r_i) + h(x) - \langle x, z\rangle \right\} \tag{41}$$

and its *Fenchel-type dual* problem

$$\sup_{v_i\in\mathcal{G}_i,\, i=1,..,m} \left\{ -\big(f^* \Box h^*\big)\left(z - \sum_{i=1}^{m} L_i^* v_i\right) - \sum_{i=1}^{m}\big(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i\rangle\big) \right\}. \tag{42}$$

Considering the maximal monotone operators

$$A = \partial f, C = \nabla h, B_i = \partial g_i \text{ and } D_i = \partial l_i,\ i = 1,..., m,$$

according to [1, Proposition 17.10, Theorem 18.15], $D_i^{-1} = \nabla l_i^*$ is a monotone and ν_i-Lipschitzian operator for $i = 1,..., m$. The monotone inclusion problem (20) reads

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in \partial f(\overline{x}) + \sum_{i=1}^{m} L_i^*((\partial g_i \Box \partial l_i)(L_i\overline{x} - r_i)) + \nabla h(\overline{x}), \tag{43}$$

while the dual inclusion problem (21) reads

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \left\{ \begin{array}{l} z - \sum_{i=1}^m L_i^* \overline{v}_i \in \partial f(x) + \nabla h(x) \\ \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i x - r_i), \ i = 1,..., m. \end{array} \right. \tag{44}$$

If $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to (43)-(44), namely,

$$z - \sum_{i=1}^m L_i^* \overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i \overline{x} - r_i), \ i = 1,..., m, \tag{45}$$

then $\overline{x}$ is an optimal solution of the problem (41), $(\overline{v}_1,..., \overline{v}_m)$ is an optimal solution of (42) and the optimal objective values of the two problems coincide. Notice that (45) is nothing else than the system of optimality condition for the primal-dual pair of convex optimization problems (41)-(42).

The assumptions made on l_i guarantees that $g_i \Box l_i \in \Gamma(\mathcal{G}_i)$ (see [1, Corollary 11.16, Proposition 12.14]) and, since $\operatorname{dom}(g_i \Box l_i) = \operatorname{dom} g_i + \operatorname{dom} l_i, \ i = 1,..., m$, one can can consider the following qualification condition of interiority-type in order to guarantee (45)

$$(r_1,..., r_m) \in \operatorname{sqri} \left(\prod_{i=1}^m (\operatorname{dom} g_i + \operatorname{dom} l_i) - \{(L_1 x,..., L_m x) : x \in \operatorname{dom} f\} \right). \tag{46}$$

The following two statements are particular instances of Algorithm 11 and Theorem 13, respectively.

Algorithm 23

`Initialization`*:*	`Choose` $\mu > 0$ `such that`
	$\mu \leq \min \left\{ \gamma^2/\eta^2, \delta_1^2/\nu_1^2,..., \delta_m^2/\nu_m^2, \sqrt{\gamma / \left(\sum_{i=1}^m \|L_i\|^2/\delta_i \right)} \right\}$,
	$\tau = \mu/(2\gamma)$, $\sigma_i = \mu/(2\delta_i)$, $i = 1, .., m$,
	$\theta \in [2/(2+\mu), 1]$ `and` $(x_0, v_{1,0},..., v_{m,0}) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$.
`For` $n \geq 0$ `set`*:*	$x_{n+1} = \operatorname{prox}_{\tau f} \left[x_n - \tau \left(\sum_{i=1}^m L_i^* v_{i,n} + \nabla h(x_n) - z \right) \right]$
	$y_n = x_{n+1} + \theta(x_{n+1} - x_n)$
	$v_{i,n+1} = \operatorname{prox}_{\sigma_i g_i^*} [v_{i,n} + \sigma_i (L_i y_n - \nabla l_i^*(v_{i,n}) - r_i)]$, $i = 1,..., m$.

Theorem 24 *Suppose that $f + h$ is γ-strongly convex for $\gamma > 0$, $g_i^* + l_i^*$ is δ_i-strongly convex for $\delta_i > 0$, $i = 1,..., m$, and the qualification condition* (46) *holds. Then there exists a unique optimal solution $\overline{x}$ to* (41)*, a unique optimal solution $(\overline{v}_1,..., \overline{v}_m)$ to* (42) *fulfilling the optimality conditions* (45) *and such that the optimal objective values of the problems* (41) *and* (42) *coincide. The sequences generated by Algorithm 23 fulfill for any $n \geq 0$*

$$\gamma \|x_{n+1} - \overline{x}\|^2 + (1 - \omega) \sum_{i=1}^m \delta_i \|v_{i,n} - \overline{v}_i\|^2 \leq$$

$$\omega^n \left(\gamma \|x_1 - \overline{x}\|^2 + \sum_{i=1}^m \delta_i \|v_{i,0} - \overline{v}_i\|^2 + \frac{\gamma}{2} \omega \|x_1 - x_0\|^2 + \mu\omega \sum_{i=1}^m \langle L_i(x_1 - x_0), v_{i,0} - \overline{v}_i \rangle \right),$$

where $0 < \omega = \frac{2(1+\theta)}{4+\mu} < 1$.

Remark 25 Let us mention that $g_i^*+l_i^*$ is δ_i-strongly convex if, for example g_i^* (or l_i^*) is δ_i-strongly convex, $i=1,...,m$. Another situation which guarantees that $g_i^*+l_i^*$ is δ_i-strongly convex is the case when g_i^* is α_i-strongly convex, l_i^* is β_i-strongly convex, where $\alpha_i,\beta_i>0$ are such that $\alpha_i+\beta_i\geq\delta_i, i=1,...,m$. Finally, according to [1, Theorem 18.15], g_i^* is α_i-strongly convex if and only if g_i is Fréchet-differentiable and ∇g_i is α_i^{-1}–Lipschitzian for $i=1,...,m$.

4 Numerical experiments

In this section we illustrate the applicability of the theoretical results in the context of two numerical experiments in image processing and pattern recognition in cluster analysis.

4.1 Image processing

In this subsection, we compare the numerical performances of Algorithm 18 with the ones of other iterative schemes recently introduced in the literature for solving an image denoising problem. To this end, we treat the nonsmooth regularized convex optimization problem

$$\inf_{x\in\mathbb{R}^k}\left\{\frac{1}{2}\left\|x-b\right\|^2+\alpha TV(x)\right\}, \tag{47}$$

where $TV:\mathbb{R}^k\to\mathbb{R}$ denotes a discrete total variation functional, $\alpha>0$ is a regularization parameter and $b\in\mathbb{R}^k$ is the observed noisy image. Notice that we consider images of size $k=M\times N$ as vectors $x\in\mathbb{R}^k$, where each pixel denoted by $x_{i,j}$, $1\leq i\leq M$, $1\leq j\leq N$, ranges in the closed interval from 0 (pure black) to 1 (pure white).

Two popular choices for the discrete total variation functional are the isotropic total variation $TV_{\text{iso}}:\mathbb{R}^k\to\mathbb{R}$,

$$\begin{aligned}TV_{\text{iso}}(x)=&\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\sqrt{(x_{i+1,j}-x_{i,j})^2+(x_{i,j+1}-x_{i,j})^2}\\&+\sum_{i=1}^{M-1}\left|x_{i+1,N}-x_{i,N}\right|+\sum_{j=1}^{N-1}\left|x_{M,j+1}-x_{M,j}\right|,\end{aligned}$$

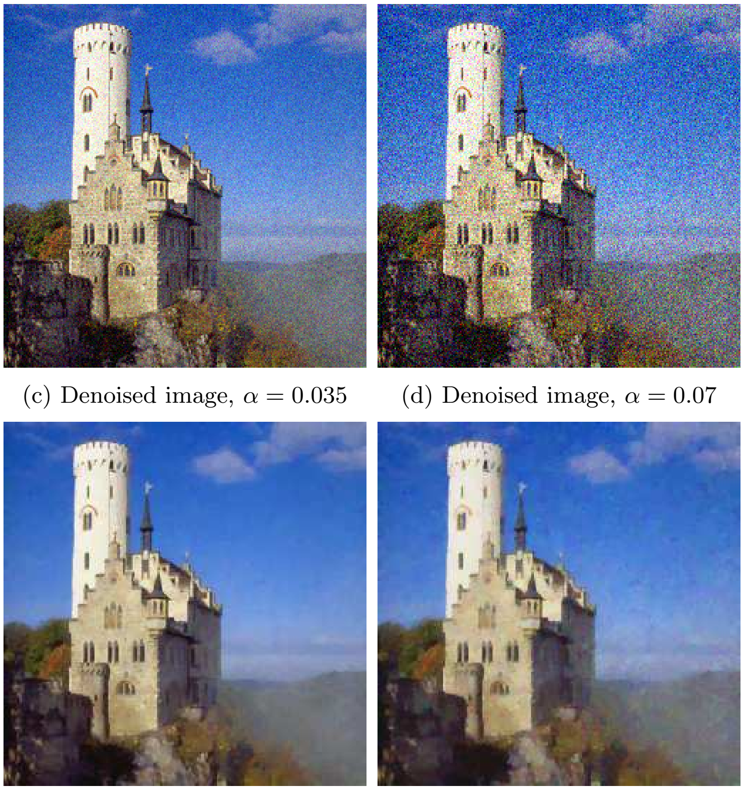
and the anisotropic total variation $TV_{\text{aniso}}:\mathbb{R}^k\to\mathbb{R}$,

$$\begin{aligned}TV_{\text{aniso}}(x)=&\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\left|x_{i+1,j}-x_{i,j}\right|+\left|x_{i,j+1}-x_{i,j}\right|\\&+\sum_{i=1}^{M-1}\left|x_{i+1,N}-x_{i,N}\right|+\sum_{j=1}^{N-1}\left|x_{M,j+1}-x_{M,j}\right|,\end{aligned}$$

where in both cases reflexive (Neumann) boundary conditions are assumed. Obviously, in both situations the qualification condition stated in Theorem 19 is fulfilled.

Denote $\mathcal{Y}=\mathbb{R}^k\times\mathbb{R}^k$ and define the linear operator $L:\mathbb{R}^k\to\mathcal{Y}$, $x_{i,j}\mapsto(L_1x_{i,j},L_2x_{i,j})$, where

$$L_1x_{i,j}=\begin{cases}x_{i+1,j}-x_{i,j}, & \text{if } i<M\\ 0, & \text{if } i=M\end{cases}\quad\text{and } L_2x_{i,j}=\begin{cases}x_{i,j+1}-x_{i,j}, & \text{if } j<N\\ 0, & \text{if } j=N\end{cases}.$$



(a) Noisy image, $\sigma = 0.06$ (b) Noisy image, $\sigma = 0.12$ (c) Denoised image, $\alpha = 0.035$ (d) Denoised image, $\alpha = 0.07$

Figure 1: The noisy images in (a) and (b) were obtained after adding white Gaussian noise with standard deviation $\sigma = 0.06$ and $\sigma = 0.12$, respectively, to the original 256×256 lichtenstein test image. The outputs of Algorithm 18 after 100 iterations when solving (47) with isotropic total variation are shown in (c) and (d), respectively.

The operator L represents a discretization of the gradient in horizontal and vertical direction. One can easily check that $\|L\|^2 \leq 8$ while for the expression of its adjoint $L^* : \mathcal{Y} \to \mathbb{R}^k$ we refer the reader to [10].

When considering the *isotropic total variation*, the problem (47) can be formulated as

$$\inf_{x \in \mathbb{R}^k} \{h(x) + g(Lx)\}, \tag{48}$$

where $h : \mathbb{R}^k \to \overline{\mathbb{R}}$, $h(x) = \frac{1}{2}\|x - b\|^2$ is 1-strongly convex with 1-Lipschitzian gradient, and $g : \mathcal{Y} \to \mathbb{R}$ is defined as $g(u, v) = \alpha\|(u, v)\|_\times$, where $\|(\cdot, \cdot)\|_\times : \mathcal{Y} \to \mathbb{R}$, $\|(u, v)\|_\times = \sum_{i=1}^M \sum_{j=1}^N \sqrt{u_{i,j}^2 + v_{i,j}^2}$, is a norm on the Hilbert space $\mathcal{Y}$. One can show (cf. [7]) that $g^*(p, q) = \delta_S(p, q)$ for every $(p, q) \in \mathcal{Y}$, where

$$S = \left\{ (p, q) \in \mathcal{Y} : \max_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} \sqrt{p_{i,j}^2 + q_{i,j}^2} \leq \alpha \right\}.$$

Moreover, by taking $(p, q) \in \mathcal{Y}$ and $\sigma > 0$, we have

$$\text{prox}_{\sigma g^*}(p, q) = P_S(p, q),$$

the projection operator $P_S : \mathcal{Y} \to S$ being defined via

$$(p_{i,j}, q_{i,j}) \mapsto \alpha \frac{(p_{i,j}, q_{i,j})}{\max\left\{\alpha, \sqrt{p_{i,j}^2 + q_{i,j}^2}\right\}}, \ 1 \leq i \leq M, \ 1 \leq j \leq N.$$

$\varepsilon = 10^{-5}$	isotropic TV		anisotropic TV	
	$\sigma = 0.06$	$\sigma = 0.12$	$\sigma = 0.06$	$\sigma = 0.12$
FB	10.55s (548)	25.78s (1335)	7.83s (517)	12.36s (829)
Algorithm 18	3.12s (177)	4.82s (275)	2.66s (202)	3.87s (290)
FBF	19.71s (698)	48.84s (1676)	15.39s (651)	24.60s (1040)
FBF Acc	3.51s (134)	5.94s (208)	3.51s (146)	4.82s (202)
AMA	19.34s (969)	45.94s (2313)	13.58s (901)	22.14s (1448)
AMA Acc	3.38s (132)	5.31s (205)	3.42s (154)	4.80s (230)
Nesterov (dual)	4.48s (146)	6.94s (230)	3.61s (172)	5.42s (249)
FISTA (dual)	3.26s (148)	5.02s (229)	3.14s (173)	4.52s (256)

Table 1: Performance evaluation for the images in Figure 1. The entries refer, respectively, to the CPU times in seconds and the number of iterations in order to attain a root-mean-square error for the primal iterates below the tolerance level of $\varepsilon = 10^{-5}$.

On the other hand, when considering the *anisotropic total variation*, the problem (47) can be formulated as

$$\inf_{x\in\mathbb{R}^k} \{h(x) + \widetilde{g}(Lx)\}, \tag{49}$$

where the function h is taken as above and $\widetilde{g} : \mathcal{Y} \to \mathbb{R}$ is defined as $\widetilde{g}(u, v) = \alpha\|(u, v)\|_1$. For every $(p, q) \in \mathcal{Y}$ we have $\widetilde{g}^*(p, q) = \delta_{[-\alpha,\alpha]^k\times[-\alpha,\alpha]^k}(p, q)$ and therefore

$$\text{prox}_{\sigma\widetilde{g}_1^*}(p, q) = P_{[-\alpha,\alpha]^k\times[-\alpha,\alpha]^k}(p, q).$$

We consider the *lichtenstein test image* of size 256 times 256 and obtain the corrupted images shown in Figure 1 by adding white Gaussian noise with standard deviation $\sigma = 0.06$ and $\sigma = 0.12$, respectively. We then solve (47) by making use of Algorithm 18 and by taking into account both instances of the discrete total variation functional. For the picture with noise level $\sigma = 0.06$, we choose the regularization parameter $\alpha = 0.035$, while, in the case when $\sigma = 0.12$, we opted for $\alpha = 0.07$. As initial choices for the parameters occuring in Algorithm 18, we let $\gamma = 0.35$, $\eta = 1$, $\lambda = \eta + 1$, $\tau_0 = 0.6\frac{2\gamma}{\eta}$, and $\sigma_0 = \frac{1}{\|L\|^2\theta_0\tau_0}$. The reconstructed images after 100 iterations for isotropic total variation are shown in Figure 1.

We compare Algorithm 18 from the point of view of the CPU time in seconds which is required in order to attain a *root-mean-square error* (RMSE) below the tolerance $\varepsilon = 10^{-5}$ with respect to the primal iterates. Therefore, Table 1 shows the achieved results where the comparison is made with the foward-backward method (FB) by Vũ in [27], the foward-backward-forward method (FBF) due to Combettes and Pesquet in [14] and its acceleration (FBF Acc) proposed in [7], the alternating minimization algorithm (AMA) from [25] and its Nesterov type (cf. [20]) acceleration (AMA Acc), as well as the FISTA (cf. [2]) and Nesterov method (cf. [21]), both operating on the dual problem.

As supported by Table 1, Algorithm 18 competes well against all these methods and provides an accelerated behavior when compared with the forward-backward method by Vũ in Theorem 2. In both of these algorithms, we made use of the ability to process the continuously differentiable function $x \mapsto \frac{1}{2}\|x - b\|^2$ via a forward evaluation of its gradient.

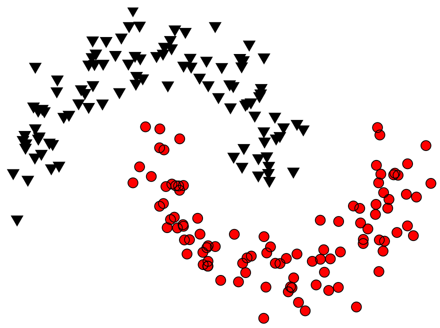


Figure 2: Clustering two interlocking half moons. The colors (resp. the shapes) show the correct affiliations.

4.2 Clustering

In cluster analysis one aims for grouping a set of points such that points within the same group are more similar to each other (usually measured via distance functions) than to points in other groups. Clustering can be formulated as a convex optimization problem (see, for instance, [12, 18, 19]). In this example, we consider the minimization problem

$$\inf_{x_i \in \mathbb{R}^n,\, i=1,\dots,m} \left\{ \frac{1}{2} \sum_{i=1}^{m} \|x_i - u_i\|^2 + \gamma \sum_{i<j} \omega_{ij} \|x_i - x_j\|_p \right\}, \tag{50}$$

where $\gamma \in \mathbb{R}_+$ is a tuning parameter, $p \in \{1, 2\}$ and $\omega_{ij} \in \mathbb{R}_+$ represent weights on the terms $\|x_i - x_j\|_p$, for $i, j = 1, \dots, m$, $i < j$. For each given point $u_i \in \mathbb{R}^n$, $i = 1, \dots, m$, the variable $x_i \in \mathbb{R}^n$ represents the associated cluster center. Since the objective function is strongly convex, there exists a unique solution to (50).

The tuning parameter $\gamma \in \mathbb{R}_+$ plays a central role within the clustering problem. Taking $\gamma = 0$, each cluster center x_i will coincide with the associated point u_i. As γ increases, the cluster centers will start to coalesce, where two points u_i, u_j are said to belong to the same cluster when $x_i = x_j$. One finally obtains a single cluster containing all points when γ becomes sufficiently large.

Moreover, the choice of the weights is important as well, since cluster centers may coalesce immediately as γ passes certain critical values. In terms of our weight selection, we use a K-nearest neighbors strategy, as proposed in [12]. Therefore, whenever $i, j = 1, \dots, m$, $i < j$, we set the weight to $\omega_{ij} = \iota_{ij}^K \exp(-\phi\|x_i - x_j\|_2^2)$, where

$$\iota_{ij}^K = \begin{cases} 1, & \text{if } j \text{ is among } i\text{'s } K\text{-nearest neighbors or vice versa,} \\ 0, & \text{otherwise.} \end{cases}$$

We consider the values $K = 10$ and $\phi = 0.5$, which are the best ones reported in [12] on a similar dataset.

Let k be the number of nonzero weights ω_{ij}. Then, one can introduce a linear operator $A : \mathbb{R}^{mn} \to \mathbb{R}^{kn}$, such that problem (50) can be equivalently written as

$$\inf_{x \in \mathbb{R}^{mn}} \{h(x) + g(Ax)\}, \tag{51}$$

the function h being 1-strongly convex and differentiable with 1-Lipschitzian gradient. Also, by taking $p \in \{1, 2\}$, the proximal points with respect to g^* admit explicit representations.

	$p = 2$, $\gamma = 5.2$		$p = 1$, $\gamma = 4$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$
FB	2.48s (1353)	5.72s (3090)	2.01s (1092)	4.05s (2226)
Algorithm 18	2.04s (1102)	4.11s (2205)	1.74s (950)	3.84s (2005)
FBF	7.67s (2123)	17.58s (4879)	6.33s (1781)	13.22s (3716)
FBF Acc	5.05s (1384)	10.27s (2801)	4.83s (1334)	9.98s (2765)
AMA	13.53s (7209)	31.09s (16630)	11.31s (6185)	23.85s (13056)
AMA Acc	3.10s (1639)	15.91s (8163)	2.51s (1392)	12.95s (7148)
Nesterov (dual)	7.85s (3811)	42.69s (21805)	7.46s (3936)	$>$ 190s ($>$ 100000)
FISTA (dual)	7.55s (4055)	51.01s (27356)	6.55s (3550)	47.81s (26069)

Table 2: Performance evaluation for the clustering problem. The entries refer to the CPU times in seconds and the number of iterations, respectively, needed in order to attain a root mean squared error for the iterates below the tolerance ε.

For our numerical tests we consider the standard dataset consisting of two interlocking half moons in $\mathbb{R}^2$, each of them being composed of 100 points (see Figure 2). The stopping criterion asks the root-mean-square error (RMSE) to be less than or equal to a given bound ε which is either $\varepsilon = 10^{-4}$ or $\varepsilon = 10^{-8}$. As tuning parameters we use $\gamma = 4$ for $p = 1$ and $\gamma = 5.2$ for $p = 2$ since both choices lead to a correct separation of the input data into the two half moons.

By taking into consideration the results given in Table 2, it shows that Algorithm 18 performs slightly better than the forward-backward (FB) method proposed in [27]. One can also see that the acceleration of the forward-backward-forward (FBF) has a positive effect on both CPU times and required iterations compared with the regular method. The alternating minimization algorithm (AMA, cf. [25]) converges slow in this example. Its Nesterov-type acceleration (cf. [20]), however, performs better. The two accelerated first-order methods FISTA (cf. [2]) and the one relying in Nesterov's scheme (cf. [21]), which are both employed on the dual problem, perform surprisingly bad in this case.

Acknowledgements. The authors are grateful to anonymous reviewers for remarks and suggestions which improved the quality of the paper.

References

[1] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-tresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. 2(1), 183–202, 2009

[3] J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge, 2010

[4] R.I. Boţ, *Conjugate Duality in Convex Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 637, Springer, Berlin Heidelberg, 2010

[5] R.I. Boţ, E.R. Csetnek, *Regularity conditions via generalized interiority notions in convex optimization: new achievements and their relation to some classical statements*, Optimization 61(1), 35–65, 2012

[6] R.I. Boţ, E.R. Csetnek, A. Heinrich, *A primal-dual splitting algorithm for fnding zeros of sums of maximally monotone operators*, SIAM Journal on Optimization 23(4),2011–2036, 2013

[7] R.I. Boţ, C. Hendrich, *Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization*, Journal of Mathematical Imaging and Vision, DOI: 10.1007/s10851-013-0486-8

[8] R.I. Boţ, C. Hendrich, *A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM Journal on Optimization 23(4), 2541–2565, 2013

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

[10] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20(1–2), 89–97, 2004

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

[12] E.C. Chi, K. Lange, *Splitting methods for convex clustering*, arXiv:1304.0499 [stat.ML], 2013

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

[14] P.L. Combettes, 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), 307–330, 2012

[15] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications 158(2), 460–479, 2013

[16] J. Douglas, H.H. Rachford, *On the numerical solution of the heat conduction problem in* 2 *and* 3 *space variables*, Transactions of the American Mathematical Society 82, 421–439, 1956

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

[18] T. Hocking, J. Vert, F. Bach, A. Joulin, *Clusterpath: an algorithm for clustering using convex fusion penalties*, In: Proceedings of the 28th International Conference on Machine Learning, Bellevue, 2011

[19] F. Lindsten, H. Ohlsson, L. Ljung, *Just relax and come clustering! A convexication of k-means clustering*, Technical report from Automatic Control, Linköpings Universitet, Report no.: LiTH-ISY-R-2992, 2011

[20] Y. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $O(1/k^2)$, Doklady AN SSSR (translated as Soviet Math. Docl.), 269, 543–547, 1983

[21] Y. Nesterov, *Smooth minimization of non-smooth functions*, Mathematical Programming 103(1), 127–152, 2005

[22] R.T. Rockafellar, *On the maximal monotonicity of subdifferential mappings*, Pacific Journal of Mathematics 33(1), 209–216, 1970

[23] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization 14(5), 877–898, 1976

[24] S. Simons, *From Hahn-Banach to Monotonicity*, Springer, Berlin, 2008

[25] P. Tseng. *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM Journal on Control and Optimization 29(1), 119–138, 1991

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

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

[28] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, Singapore, 2002

[29] M. Zhu, T. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, Cam Reports 08-34 UCLA, Center for Applied Mathematics, 2008