Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization

Radu Ioan Boţ * Christopher Hendrich †

December 20, 2013

Abstract. In this paper we investigate the convergence behavior of a primal-dual splitting method for solving monotone inclusions involving mixtures of composite, Lipschitzian and parallel sum type operators proposed by Combettes and Pesquet in [10]. Firstly, in the particular case of convex minimization problems, we derive convergence rates for the partial primal-dual gap function associated to a primal-dual pair of optimization problems by making use of conjugate duality techniques. Secondly, we propose for the general monotone inclusion problem two new schemes which accelerate the sequences of primal and/or dual iterates, provided strong monotonicity assumptions for some of the involved operators are fulfilled. Finally, we apply the theoretical achievements in the context of different types of image restoration problems solved via total variation regularization.

Keywords. splitting method, Fenchel duality, convergence statements, image processing

AMS subject classification. 90C25, 90C46, 47A52

1 Introduction

1.1 Review of previous work

Motivated by applications in fields like signal and image processing, location theory and supervised machine learning, the last few years have shown a rising interest in solving structured nondifferentiable convex optimization problems within the framework of the theory of conjugate functions. As this gives rise via convex duality and optimality statements to the solving of monotone inclusions involving mixtures of composite, single-valued cocoercive and/or Lipschitzian and parallel sum type operators, the focus was put on providing easily implementable numerical schemes for the latter. In this sense, one of the major aim was to avoid asking for the calculation of the resolvents of the composites with linear continuous operators and of the parallel sum types operators, for which in general no exact formulae are available, rather then to evaluate each maximally

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

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

monotone operator individually. Another important aim was to include the single-valued cocoercive and/or Lipschitzian operators via forward evaluations.

As the standard splitting approaches, like the Forward-Backward algorithm (see [1]), Tseng's Forward-Backward-Forward algorithm (see [13]) and the Douglas–Rachford algorithm (see [1]), proved to have considerable limitations in this context, a first fruitful idea in this sense, was proposed by Combettes and Pesquet in [10], itself being an extension of the algorithmic scheme from [7] obtained by allowing also Lipschitzian monotone operators and parallel sums in the problem formulation. In the mentioned works, by means of a primal-dual reformulation in an appropriate product space, the monotone inclusion problem is reduced to the one of finding the zeros of the sum of a Lipschitzian monotone operator with a maximally monotone operator. The latter is solved by using an error-tolerant version of Tseng's algorithm which has forward-backward-forward characteristics and allows to access the monotone Lipschitzian operators via explicit forward steps, while set-valued maximally monotone operators are processed via their resolvents. A notable advantage of this method is given by both its highly parallelizable character, most of its steps could be executed independently, and by the fact that allows to process maximal monotone operators and linear bounded operators separately, whenever they occur in the form of precompositions in the problem formulation.

This idea was further employed by Vũ in [15] in the context of solving highly structured monotone inclusions, as well, whereby instead of monotone Lipschitzian operators, cocoercive operators were used and, consequently, instead of Tseng's splitting, the forward-backward splitting method has been used. The popular primal-dual method due to Chambolle and Pock described and analyzed in [9, Algorithm 1] and its extension proposed by Condat in [11] are particular instances of Vũ's algorithm. For other recently introduced algorithms for solving monotone inclusions relying on the primal-dual approach we refer the reader to [4–6].

The aim of this paper is to investigate the convergence behavior of a primal-dual splitting method from [10] from two different points of view, namely, by deriving convergence rates for the sequence of objective function values in the particular case of convex minimization problems and by proposing two new schemes which accelerate the sequences of primal and/or dual iterates. The theoretical achievements are then applied in the context of different types of image restoration problems solved via total variation regularization.

1.2 Preliminary notions and problem formulation

We introduce as follows some preliminary notions and results which are needed throughout the paper and formulate the monotone inclusion problem under investigation. We are considering the real Hilbert spaces $\mathcal{H}$ and $\mathcal{G}_i$, $i = 1, \dots, m$, endowed with the *inner product* $\langle \cdot, \cdot \rangle$ and associated *norm* $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, for which we use the same notation, respectively, as there is no risk of confusion. The symbols $\rightharpoonup$ and $\rightarrow$ denote weak and strong convergence, respectively. By $\mathbb{R}_{++}$ we denote the set of strictly positive real numbers, while the *indicator function* of a set $C \subseteq \mathcal{H}$ is $\delta_C : \mathcal{H} \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, defined by $\delta_C(x) = 0$ for $x \in C$ and $\delta_C(x) = +\infty$, otherwise. For a function $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ we denote by $\operatorname{dom} f := \{x \in \mathcal{H} : f(x) < +\infty\}$ its *effective domain* and call f *proper* if $\operatorname{dom} f \neq \varnothing$ and $f(x) > -\infty$ for all $x \in \mathcal{H}$. Let be

$$\Gamma(\mathcal{H}) := \{f : \mathcal{H} \rightarrow \overline{\mathbb{R}} : f \text{ is proper, convex and lower semicontinuous}\}.$$

The *conjugate function* of f is $f^* : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, $f^*(p) = \sup\{\langle p, x \rangle - f(x) : x \in \mathcal{H}\}$ for all $p \in \mathcal{H}$ and, if $f \in \Gamma(\mathcal{H})$, then $f^* \in \Gamma(\mathcal{H})$, as well. The *(convex) subdifferential* of

$f : \mathcal{H} \to \overline{\mathbb{R}}$ at $x \in \mathcal{H}$ is the set $\partial f(x) = \{p \in \mathcal{H} : f(y) - f(x) \geq \langle p, y - x\rangle \ \forall y \in \mathcal{H}\}$, if $f(x) \in \mathbb{R}$, and is taken to be the empty set, otherwise. For a linear continuous operator $L_i : \mathcal{H} \to \mathcal{G}_i$, the operator $L_i^* : \mathcal{G}_i \to \mathcal{H}$, defined via $\langle L_i x, y\rangle = \langle x, L_i^* y\rangle$ for all $x \in \mathcal{H}$ and all $y \in \mathcal{G}_i$, denotes its *adjoint operator*, for $i \in \{1, \ldots, m\}$.

Having two functions $f, g : \mathcal{H} \to \overline{\mathbb{R}}$, their *infimal convolution* is defined by $f \,\Box\, g : \mathcal{H} \to \overline{\mathbb{R}}$, $(f \,\Box\, g)(x) = \inf_{y \in \mathcal{H}} \{f(y) + g(x - y)\}$ for all $x \in \mathcal{H}$, being a convex function when f and g are convex.

Let $M : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. We denote by $\operatorname{gra} M = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Mx\}$ its *graph* and by $\operatorname{ran} M = \{u \in \mathcal{H} : \exists x \in \mathcal{H},\ u \in Mx\}$ its *range*. The *inverse operator* of M is defined as $M^{-1} : \mathcal{H} \to 2^{\mathcal{H}}, M^{-1}(u) = \{x \in H : u \in Mx\}$. The operator M is called *monotone* if $\langle x - y, u - v\rangle \geq 0$ for all $(x, u), (y, v) \in \operatorname{gra} M$ and it is called *maximally monotone* if there exists no monotone operator $M' : \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} M'$ properly contains $\operatorname{gra} M$. The operator M is called ρ-strongly monotone, for $\rho \in \mathbb{R}_{++}$, if $M - \rho\mathrm{Id}$ is monotone, i. e. $\langle x - y, u - v\rangle \geq \rho\|x - y\|^2$ for all $(x, u), (y, v) \in \operatorname{gra} M$, where Id denotes the identity on $\mathcal{H}$. The operator $M : \mathcal{H} \to \mathcal{H}$ is called ν-Lipschitzian for $\nu \in \mathbb{R}_{++}$ if it is single-valued and it fulfills $\|Mx - My\| \leq \nu\|x - y\|$ for all $x, y \in \mathcal{H}$.

The resolvent of a set-valued operator $M : \mathcal{H} \to 2^{\mathcal{H}}$ is $J_M : \mathcal{H} \to 2^{\mathcal{H}}, J_M = (\mathrm{Id} + M)^{-1}$. When M is maximally monotone, the resolvent is a single-valued, 1-Lipschitzian and maximal monotone operator. Moreover, when $f \in \Gamma(\mathcal{H})$ and $\gamma \in \mathbb{R}_{++}$, $\partial(\gamma f)$ is maximally monotone (cf. [16, Theorem 3.2.8]) and it holds $J_{\gamma\partial f} = (\mathrm{Id} + \gamma\partial f)^{-1} = \mathrm{Prox}_{\gamma f}$. Here, $\mathrm{Prox}_{\gamma f}(x)$ denotes the *proximal point* of parameter γ of f at $x \in \mathcal{H}$ and it represents the unique optimal solution of the optimization problem

$$\inf_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}. \tag{1.1}$$

For a nonempty, convex and closed set $C \subseteq \mathcal{H}$ and $\gamma \in \mathbb{R}_{++}$ we have $\mathrm{Prox}_{\gamma\delta_C} = \mathcal{P}_C$, where $\mathcal{P}_C : \mathcal{H} \to C$, $\mathcal{P}_C(x) = \arg\min_{z \in C} \|x - z\|$, denotes the *projection operator* on C.

Finally, the *parallel sum* of two set-valued operators $M_1, M_2 : \mathcal{H} \to 2^{\mathcal{H}}$ is defined as

$$M_1 \,\Box\, M_2 : \mathcal{H} \to 2^{\mathcal{H}}, M_1 \,\Box\, M_2 = \left(M_1^{-1} + M_2^{-1}\right)^{-1}.$$

We come now to the formulation of the monotone inclusion problem which we aim to investigate throughout this paper (see [10]).

Problem 1.1. Consider the real Hilbert spaces $\mathcal{H}$ and $\mathcal{G}_i, i = 1, ..., m$, $A : \mathcal{H} \to 2^{\mathcal{H}}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ a monotone and μ-Lipschitzian operator for some $\mu \in \mathbb{R}_{++}$. Furthermore, let $z \in \mathcal{H}$ and for every $i \in \{1, \ldots, m\}$, let $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be maximally monotone operators, let $D_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be monotone operators such that D_i^{-1} is ν_i-Lipschitzian for some $\nu_i \in \mathbb{R}_{++}$, and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^* \left((B_i \,\Box\, D_i)(L_i\overline{x} - r_i)\right) + C\overline{x}, \tag{1.2}$$

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1, \ldots, \overline{v}_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \left\{ \begin{array}{l} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in (B_i \,\Box\, D_i)(L_i x - r_i), i = 1, \ldots, m. \end{array} \right. \tag{1.3}$$

Throughout this paper we denote by $\boldsymbol{\mathcal{G}} := \mathcal{G}_1 \times ... \times \mathcal{G}_m$ the Hilbert space equipped with the inner product

$$\langle (p_1, \dots, p_m), (q_1, \dots, q_m) \rangle = \sum_{i=1}^{m} \langle p_i, q_i \rangle \ \forall (p_1, \dots, p_m) \ \forall (q_1, \dots, q_m) \in \boldsymbol{\mathcal{G}}$$

and the associated norm $\|(p_1, \dots, p_m)\| = \sqrt{\sum_{i=1}^m \|p_i\|^2}$ for all $(p_1, \dots, p_m) \in \boldsymbol{\mathcal{G}}$. We introduce also the nonzero linear continuous operator $\boldsymbol{L} : \mathcal{H} \to \boldsymbol{\mathcal{G}}$, $\boldsymbol{L}x = (L_1 x, \dots, L_m x)$, its adjoint being $\boldsymbol{L}^* : \boldsymbol{\mathcal{G}} \to \mathcal{H}$, $\boldsymbol{L}^*\boldsymbol{v} = \sum_{i=1}^m L_i^* v_i$.

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

$$z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in A\overline{x} + C\overline{x} \text{ and } \overline{v}_i \in (B_i \,\Box\, D_i)(L_i \overline{x} - r_i),\ i = 1, \dots, m. \tag{1.4}$$

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

$$\overline{x} \text{ solves (1.2)} \Leftrightarrow z - \sum_{i=1}^{m} L_i^* (B_i \,\Box\, D_i)(L_i \overline{x} - r_i) \in A\overline{x} + C\overline{x} \Leftrightarrow$$

$$\exists\, \overline{v}_1 \in \mathcal{G}_1, \dots, \overline{v}_m \in \mathcal{G}_m \text{ such that } \begin{cases} z - \sum_{i=1}^m L_i^* \overline{v}_i \in A\overline{x} + C\overline{x}, \\ \overline{v}_i \in (B_i \,\Box\, D_i)(L_i \overline{x} - r_i),\ i = 1, \dots, m. \end{cases}$$

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

1.3 Contributions

The investigations we make in this paper have as starting point the primal-dual algorithm for solving Problem 1.1 given in [10, Theorem 3.1]. Firstly, we consider Problem 1.1 in its particular formulation as a primal-dual pair of convex optimization problems, approach which relies on the fact that the subdifferential of a proper, convex and lower semicontinuous function is maximally monotone. By assuming that the sequence of step sizes in the algorithm in [10, Theorem 3.1] is nondecreasing and by making use of some ideas provided in [9] we prove that the convergence rate of the partial primal-dual gap function associated to the primal-dual pair of optimization problems at a primal-dual pair of generated iterates in ergodic sense is of $\mathcal{O}(\frac{1}{n})$, where $n \in \mathbb{N}$ is the number of passed iterations. From here we are able to derive under some appropriate assumptions the same rate of convergence, again in ergodic sense, for the sequence of primal objective function values on the iterates generated by the numerical scheme.

Further, in Section 3, we provide for the general monotone inclusion problem, as given in Problem 1.1, two new acceleration schemes which generate under strong monotonicity assumptions sequences of primal and/or dual iterates that converge with improved convergence properties. To this end we use the fruitful idea of variable step sizes that have been first utilized in [17] and then shown in [9] to yield an accelerated algorithm in the case of convex optimization problems.

The feasibility and the functionality of the proposed methods are explicitly shown in Section 4 by means of numerical experiments in the context of solving image denoising, image deblurring and image inpainting problems via total variation regularization. When dealing with image denoising comparisons to popular algorithms, that proved to have good performances in this context, are made.

2 Convex minimization problems

The aim of this section is to provide a rate of convergence for the sequence of the values of the objective function at the iterates generated by a slight modification of the algorithm in [10, Theorem 3.1] when employed in the solving a convex minimization problem and its conjugate dual. The primal-dual pair under investigation is described in the following.

Problem 2.1. Consider the real Hilbert spaces $\mathcal{H}$ and $\mathcal{G}_i, i = 1,...,m$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with μ-Lipschitzian gradient for some $\mu \in \mathbb{R}_{++}$. Furthermore, let $z \in \mathcal{H}$ and for every $i \in \{1,\dots,m\}$, let $r_i \in \mathcal{G}_i$, $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that l_i is ν_i^{-1}-strongly convex for some $\nu_i \in \mathbb{R}_{++}$, and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. We consider the convex minimization problem

$$(P) \quad \inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \,\square\, l_i)(L_i x - r_i) + h(x) - \langle x, z\rangle \right\} \tag{2.1}$$

and its dual problem

$$(D) \quad \sup_{(v_i,\dots,v_m)\in\mathcal{G}_1\times\ldots\times\mathcal{G}_m} \left\{ -\left(f^* \,\square\, h^*\right)\left(z - \sum_{i=1}^{m} L_i^* v_i\right) - \sum_{i=1}^{m} \left(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i\rangle\right) \right\}. \tag{2.2}$$

In order to investigate the primal-dual pair (2.1)-(2.2) in the context of Problem 2.1, one has to take

$$A = \partial f, \ C = \nabla h, \ \text{and, for } i = 1,\dots,m, \ B_i = \partial g_i \text{ and } D_i = \partial l_i. \tag{2.3}$$

Then A and $B_i, i = 1,...,m$ are maximal monotone, C is monotone, by [1, Proposition 17.10], and $D_i^{-1} = \nabla l_i^*$ is monotone and ν_i-Lipschitz continuous for $i = 1,\dots,m$, according to [1, Proposition 17.10, Theorem 18.15 and Corollary 16.24]. One can easily see that (see, for instance, [10, Theorem 4.2]) whenever $(\overline{x}, \overline{v}_1,\dots,\overline{v}_m) \in \mathcal{H}\times\boldsymbol{\mathcal{G}}$ is a primal-dual solution to Problem 1.1, with the above choice of the involved operators, $\overline{x}$ is an optimal solution to (P), $(\overline{v}_1,\dots,\overline{v}_m)$ is an optimal solution to (D) and for (P)-(D) strong duality holds, thus the optimal objective values of the two problems coincide. On the other hand, when $\overline{x}$ is an optimal solution to (P) and a qualification condition, like (see, for instance, [3, 10])

$$\bigcup_{\lambda\geq 0} \lambda\{(L_1x - y_1 - r_1, ..., L_mx - y_m - r_m) : x \in \operatorname{dom} f, y_i \in \operatorname{dom} g_i + \operatorname{dom} l_i, i = 1,..,m\} \text{ is}$$
$$\text{a closed linear suspace,}$$

is fulfilled, then there exists $(\overline{v}_1,\dots,\overline{v}_m)$, an optimal solution to (D), such that $(\overline{x}, \overline{v}_1,\dots,\overline{v}_m) \in \mathcal{H}\times\boldsymbol{\mathcal{G}}$ is a primal-dual solution to Problem 1.1 in the particular formulation given by (2.3).

The primal-dual pair in Problem 2.1 captures various types of optimization problems. One such particular instance is formulated as follows and we refer for more examples to [10].

Example 2.1. In Problem 2.1 take $z = 0$, let $l_i : \mathcal{G}_i \to \overline{\mathbb{R}}$, $l_i = \delta_{\{0\}}$ and $r_i = 0$ for $i = 1,\dots,m$, and set $h : \mathcal{H} \to \mathbb{R}$, $h(x) = 0$ for all $x \in \mathcal{H}$. Then (2.1) reduces to

$$(P) \quad \inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} g_i(L_i x) \right\},$$

while the dual problem (2.2) becomes

$$(D) \qquad \sup_{(v_i,\dots,v_m)\in\mathcal{G}_1\times\ldots\times\mathcal{G}_m} \left\{-f^*\left(-\sum_{i=1}^m L_i^* v_i\right) - \sum_{i=1}^m g_i^*(v_i)\right\}.$$

In order to simplify the upcoming formulations and calculations we introduce the following more compact notations. With respect to Problem 2.1, let $F : \mathcal{H} \to \overline{\mathbb{R}}$, $F(x) = f(x) + h(x) - \langle x, z\rangle$. Then $\operatorname{dom} F = \operatorname{dom} f$ and its conjugate $F^* : \mathcal{H} \to \overline{\mathbb{R}}$ is given by $F^*(p) = (f + h)^*(z + p) = (f^* \Box h^*)(z + p)$, since $\operatorname{dom} h = \mathcal{H}$. Further, we set

$$\boldsymbol{v} = (v_1,\dots,v_m), \quad \overline{\boldsymbol{v}} = (\overline{v}_1,\dots,\overline{v}_m), \quad \boldsymbol{p}_{2,n} = (p_{2,1,n},\dots,p_{2,m,n}), \quad \boldsymbol{r} = (r_1,\dots,r_m).$$

We define the function $G : \boldsymbol{\mathcal{G}} \to \overline{\mathbb{R}}$, $G(\boldsymbol{y}) = \sum_{i=1}^m (g_i \Box l_i)(y_i)$ and observe that its conjugate $G^* : \boldsymbol{\mathcal{G}} \to \overline{\mathbb{R}}$ is given by $G^*(\boldsymbol{v}) = \sum_{i=1}^m (g_i \Box l_i)^*(v_i) = \sum_{i=1}^m (g_i^* + l_i^*)(v_i)$. Notice that, as $l_i^*, i = 1,\dots,m$, has full domain (cf. [1, Theorem 18.15]), we get

$$\operatorname{dom} G^* = (\operatorname{dom} g_1^* \cap \operatorname{dom} l_1^*) \times \ldots \times (\operatorname{dom} g_m^* \cap \operatorname{dom} l_m^*) = \operatorname{dom} g_1^* \times \ldots \times \operatorname{dom} g_m^*, \tag{2.4}$$

The primal and the dual optimization problems given in Problem 2.1 can be equivalently represented as

$$(P) \qquad \inf_{x\in\mathcal{H}} \{F(x) + G(\boldsymbol{L}x - \boldsymbol{r})\},$$

and, respectively,

$$(D) \qquad \sup_{\boldsymbol{v}\in\boldsymbol{\mathcal{G}}} \{-F^*(-\boldsymbol{L}^*\boldsymbol{v}) - G^*(\boldsymbol{v}) - \langle \boldsymbol{v}, \boldsymbol{r}\rangle\}.$$

Then $\overline{x} \in \mathcal{H}$ solves (P), $\overline{\boldsymbol{v}} \in \boldsymbol{\mathcal{G}}$ solves (D) and for (P)-(D) strong duality holds if and only if (cf. [3])

$$-\boldsymbol{L}^*\overline{\boldsymbol{v}} \in \partial F(\overline{x}) \text{ and } \boldsymbol{L}\overline{x} - \boldsymbol{r} \in \partial G^*(\overline{\boldsymbol{v}}). \tag{2.5}$$

Let us mention also that for $\overline{x} \in \mathcal{H}$ and $\overline{\boldsymbol{v}} \in \boldsymbol{\mathcal{G}}$ fulfilling (2.5) it holds

$$[\langle \boldsymbol{L}x - \boldsymbol{r}, \overline{\boldsymbol{v}}\rangle + F(x) - G^*(\overline{\boldsymbol{v}})] - [\langle \boldsymbol{L}\overline{x} - \boldsymbol{r}, \boldsymbol{v}\rangle + F(\overline{x}) - G^*(\boldsymbol{v})] \geq 0 \ \forall x \in \mathcal{H} \ \forall \boldsymbol{v} \in \boldsymbol{\mathcal{G}}.$$

For given sets $B_1 \subseteq \mathcal{H}$ and $B_2 \subseteq \boldsymbol{\mathcal{G}}$ we introduce the so-called *primal-dual gap function restricted to* $B_1 \times B_2$

$$\begin{aligned}\mathcal{G}_{B_1\times B_2}(x,\boldsymbol{v}) = &\sup_{\tilde{\boldsymbol{v}}\in B_2} \{\langle \boldsymbol{L}x - \boldsymbol{r}, \tilde{\boldsymbol{v}}\rangle + F(x) - G^*(\tilde{\boldsymbol{v}})\} \\ &- \inf_{\tilde{x}\in B_1} \{\langle \boldsymbol{L}\tilde{x} - \boldsymbol{r}, \boldsymbol{v}\rangle + F(\tilde{x}) - G^*(\boldsymbol{v})\}.\end{aligned} \tag{2.6}$$

We consider the following algorithm for solving (P)-(D), which differs from the primal-dual one given by Combettes and Pesquet in [10, Theorem 3.1] by the fact that we are asking the sequence $(\gamma_n)_{n\geq 0} \subseteq \mathbb{R}_{++}$ to be nondecreasing.

Algorithm 2.1. Let $x_0 \in \mathcal{H}$ and $(v_{1,0},\dots,v_{m,0}) \in \boldsymbol{\mathcal{G}}$, set

$$\beta = \max\{\mu, \nu_1,\dots,\nu_m\} + \sqrt{\sum_{i=1}^n \|L_i\|^2},$$

choose $\varepsilon \in \left(0, \frac{1}{\beta+1}\right)$ and $(\gamma_n)_{n\geq 0}$ a nondecreasing sequence in $\left[\varepsilon, \frac{1-\varepsilon}{\beta}\right]$ and set

$$(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} p_{1,n} = \operatorname{Prox}_{\gamma_n f}\left(x_n - \gamma_n\left(\nabla h(x_n) + \sum_{i=1}^m L_i^* v_{i,n} - z\right)\right) \\ \text{For } i = 1, \ldots, m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = \operatorname{Prox}_{\gamma_n g_i^*}\left(v_{i,n} + \gamma_n(L_i x_n - \nabla l_i^*(v_{i,n}) - r_i)\right) \\ v_{i,n+1} = \gamma_n L_i(p_{1,n} - x_n) + \gamma_n(\nabla l_i^*(v_{i,n}) - \nabla l_i^*(p_{2,i,n})) + p_{2,i,n} \end{array} \right. \\ x_{n+1} = \gamma_n \sum_{i=1}^m L_i^*(v_{i,n} - p_{2,i,n}) + \gamma_n(\nabla h(x_n) - \nabla h(p_{1,n})) + p_{1,n}. \end{array} \right. \tag{2.7}$$

The following theorem is formulated in the spirit of [9, Theorem 1], however, the techniques of the proof used are adjusted to the forward-backward-forward structure of Algorithm 2.1.

Theorem 2.1. *For Problem 2.1 suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^m L_i^*\left(\partial g_i \,\Box\, \partial l_i\right)(L_i \cdot - r_i) + \nabla h\right).$$

Then there exists an optimal solution $\overline{x} \in \mathcal{H}$ to (P) and an optimal solution $(\overline{v}_1, \ldots, \overline{v}_m) \in \boldsymbol{\mathcal{G}}$ to (D), such that the following holds for the sequences generated by Algorithm 2.1:

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

(b) $x_n \rightharpoonup \overline{x}$, $p_{1,n} \rightharpoonup \overline{x}$ *and* $v_{i,n} \rightharpoonup \overline{v}_i$, $p_{2,i,n} \rightharpoonup \overline{v}_i$ $\forall i \in \{1, \ldots, m\}$.

(c) For $n \geq 0$ it holds

$$\frac{\|x_n - \overline{x}\|^2}{2\gamma_n} + \sum_{i=1}^m \frac{\|v_{i,n} - \overline{v}_i\|^2}{2\gamma_n} \leq \frac{\|x_0 - \overline{x}\|^2}{2\gamma_0} + \sum_{i=1}^m \frac{\|v_{i,0} - \overline{v}_i\|^2}{2\gamma_0}.$$

(d) If $B_1 \subseteq \mathcal{H}$ and $B_2 \subseteq \boldsymbol{\mathcal{G}}$ are bounded, then for $x^N := \frac{1}{N}\sum_{n=0}^{N-1} p_{1,n}$ and $v_i^N := \frac{1}{N}\sum_{n=0}^{N-1} p_{2,i,n}$, $i = 1, \ldots, m$, the primal-dual gap has the upper bound

$$\mathcal{G}_{B_1 \times B_2}(x^N, v_1^N, \ldots, v_m^N) \leq \frac{C(B_1, B_2)}{N}, \tag{2.8}$$

where

$$C(B_1, B_2) = \sup_{(x, v_1, \ldots, v_m) \in B_1 \times B_2} \left\{ \frac{\|x_0 - x\|^2}{2\gamma_0} + \sum_{i=1}^m \frac{\|v_{i,0} - v_i\|^2}{2\gamma_0} \right\}.$$

(e) The sequence $(x^N, v_1^N, \ldots, v_m^N)$ converges weakly to $(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m)$.

Proof. Theorem 4.2 in [10] guarantees the existence of an optimal solution $\overline{x} \in \mathcal{H}$ to (2.1) and of an optimal solution $(\overline{v}_1, \ldots, \overline{v}_m) \in \boldsymbol{\mathcal{G}}$ to (2.2) such that strong duality holds, $x_n \rightharpoonup \overline{x}$, $p_{1,n} \rightharpoonup \overline{x}$, as well as $v_{i,n} \rightharpoonup \overline{v}_i$ and $p_{2,i,n} \rightharpoonup \overline{v}_i$ for $i = 1, \ldots, m$, when n converges to $+\infty$. Hence *(a)* and *(b)* are true. Thus, the solutions $\overline{x}$ and $\overline{\boldsymbol{v}} = (\overline{v}_1, \ldots, \overline{v}_m)$ fulfill (2.5).

Regarding the sequences $(p_{1,n})_{n\geq 0}$ and $(p_{2,i,n})_{n\geq 0}$, $i = 1, \ldots, m$, generated in Algorithm 2.1 we have for every $n \geq 0$

$$\begin{aligned} p_{1,n} = (\mathrm{Id} + \gamma_n \partial f)^{-1}\left(x_n - \gamma_n\left(\nabla h(x_n) + L^* v_n - z\right)\right) \\ \Leftrightarrow \frac{x_n - p_{1,n}}{\gamma_n} - \nabla h(x_n) - L^* v_n + z \in \partial f(p_{1,n}) \end{aligned}$$

and, for $i = 1, ..., m$,

$$p_{2,i,n} = (\mathrm{Id} + \gamma_n \partial g_i^*)^{-1} (v_{i,n} + \gamma_n(L_i x_n - \nabla l_i^*(v_{i,n}) - r_i))$$
$$\Leftrightarrow \frac{v_{i,n} - p_{2,i,n}}{\gamma_n} + L_i x_n - \nabla l_i^*(v_{i,n}) - r_i \in \partial g_i^*(p_{2,i,n}).$$

In other words, it holds for every $n \geq 0$

$$f(x) \geq f(p_{1,n}) + \left\langle \frac{x_n - p_{1,n}}{\gamma_n} - \nabla h(x_n) - L^* v_n + z, x - p_{1,n} \right\rangle \forall x \in \mathcal{H} \tag{2.9}$$

and, for $i = 1, \ldots, m$,

$$g_i^*(v_i) \geq g_i^*(p_{2,i,n}) + \left\langle \frac{v_{i,n} - p_{2,i,n}}{\gamma_n} + L_i x_n - \nabla l_i^*(v_{i,n}) - r_i, v_i - p_{2,i,n} \right\rangle \forall v_i \in \mathcal{G}_i. \tag{2.10}$$

In addition to that, using that h and $l_i^*, i = 1, ..., m$, are convex and differentiable, it holds for every $n \geq 0$

$$h(x) \geq h(p_{1,n}) + \langle \nabla h(p_{1,n}), x - p_{1,n} \rangle \ \forall x \in \mathcal{H} \tag{2.11}$$

and, for $i = 1, \ldots, m$,

$$l_i^*(v_i) \geq l_i^*(p_{2,i,n}) + \langle \nabla l_i^*(p_{2,i,n}), v_i - p_{2,i,n} \rangle \ \forall v_i \in \mathcal{G}_i. \tag{2.12}$$

Consider arbitrary $x \in \mathcal{H}$ and $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m) \in \boldsymbol{\mathcal{G}}$. Since

$$\left\langle \frac{x_n - p_{1,n}}{\gamma_n}, x - p_{1,n} \right\rangle = \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|x - p_{1,n}\|^2}{2\gamma_n} - \frac{\|x_n - x\|^2}{2\gamma_n}$$
$$\left\langle \frac{v_{i,n} - p_{2,i,n}}{\gamma_n}, v_i - p_{2,i,n} \right\rangle = \frac{\|v_{i,n} - p_{2,i,n}\|^2}{2\gamma_n} + \frac{\|v_i - p_{2,i,n}\|^2}{2\gamma_n} - \frac{\|v_{i,n} - v_i\|^2}{2\gamma_n}, i = 1, ..., m,$$

we obtain for every $n \geq 0$, by using the more compact notation of the elements in $\boldsymbol{\mathcal{G}}$ and by summing up the inequalities (2.9)–(2.12),

$$\frac{\|x_n - x\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \boldsymbol{v}\|^2}{2\gamma_n} \geq \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|x - p_{1,n}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v} - \boldsymbol{p}_{2,n}\|^2}{2\gamma_n}$$
$$+ \sum_{i=1}^{m} \langle L_i x_n + \nabla l_i^*(p_{2,i,n}) - \nabla l_i^*(v_{i,n}) - r_i, v_i - p_{2,i,n} \rangle - \sum_{i=1}^{m} (g_i^* + l_i^*)(v_i) + (f + h)(p_{1,n})$$
$$+ \langle \nabla h(p_{1,n}) - \nabla h(x_n) - L^* v_n + z, x - p_{1,n} \rangle - \left[\sum_{i=1}^{m} -(g_i^* + l_i^*)(p_{2,i,n}) + (f + h)(x) \right].$$

Further, using again the update rules in Algorithm 2.1 and the equations

$$\left\langle \frac{p_{1,n} - x_{n+1}}{\gamma_n}, x - p_{1,n} \right\rangle = \frac{\|x_{n+1} - x\|^2}{2\gamma_n} - \frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} - \frac{\|x - p_{1,n}\|^2}{2\gamma_n}$$

and, for $i = 1, ..., m$,

$$\left\langle \frac{p_{2,i,n} - v_{i,n+1}}{\gamma_n}, v_i - p_{2,i,n} \right\rangle = \frac{\|v_{i,n+1} - v_i\|^2}{2\gamma_n} - \frac{\|v_{i,n+1} - p_{2,i,n}\|^2}{2\gamma_n} - \frac{\|v_i - p_{2,i,n}\|^2}{2\gamma_n},$$

we obtain for every $n \geq 0$

$$\begin{aligned}
\frac{\|x_n - x\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \boldsymbol{v}\|^2}{2\gamma_n} \geq & \frac{\|x_{n+1} - x\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{v}\|^2}{2\gamma_n} + \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2}{2\gamma_n} \\
& - \frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} - \frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{p}_{2,n}\|^2}{2\gamma_n} + [\langle \boldsymbol{L}p_{1,n} - \boldsymbol{r}, \boldsymbol{v} \rangle - G^*(\boldsymbol{v}) + F(p_{1,n})] \\
& - \left[\left\langle \boldsymbol{L}x - \boldsymbol{r}, \boldsymbol{p}_{2,n} \right\rangle - G^*(\boldsymbol{p}_{2,n}) + F(x) \right]. \qquad (2.13)
\end{aligned}$$

Further, we equip the Hilbert space $\boldsymbol{\mathcal{H}} = \mathcal{H} \times \boldsymbol{\mathcal{G}}$ with the inner product

$$\langle (y, \boldsymbol{p}), (z, \boldsymbol{q}) \rangle = \langle y, z \rangle + \langle \boldsymbol{p}, \boldsymbol{q} \rangle \ \forall (y, \boldsymbol{p}), (z, \boldsymbol{q}) \in \mathcal{H} \times \boldsymbol{\mathcal{G}} \qquad (2.14)$$

and the associated norm $\|(y, \boldsymbol{p})\| = \sqrt{\|y\|^2 + \|\boldsymbol{p}\|^2}$ for every $(y, \boldsymbol{p}) \in \mathcal{H} \times \boldsymbol{\mathcal{G}}$. For every $n \geq 0$ it holds

$$\frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{p}_{2,n}\|^2}{2\gamma_n} = \frac{\|(x_{n+1}, \boldsymbol{v}_{n+1}) - (p_{1,n}, \boldsymbol{p}_{2,n})\|^2}{2\gamma_n}$$

and, consequently, by making use of the Lipschitz continuity of ∇h and ∇l_i^*, $i = 1, \ldots, m$, it shows that

$$\begin{aligned}
& \|(x_{n+1}, \boldsymbol{v}_{n+1}) - (p_{1,n}, \boldsymbol{p}_{2,n})\| \\
& = \gamma_n \|(\boldsymbol{L}^*(\boldsymbol{v}_n - \boldsymbol{p}_{2,n}), L_1(p_{1,n} - x_n), \ldots, L_m(p_{1,n} - x_n)) \\
& \qquad + (\nabla h(x_n) - \nabla h(p_{1,n}), \nabla l_1^*(v_{1,n}) - \nabla l_1^*(p_{2,1,n}), \ldots, \nabla l_m^*(v_{m,n}) - \nabla l_1^*(p_{2,m,n}))\| \\
& \leq \gamma_n \|(\boldsymbol{L}^*(\boldsymbol{v}_n - \boldsymbol{p}_{2,n}), L_1(p_{1,n} - x_n), \ldots, L_m(p_{1,n} - x_n))\| \\
& \quad + \gamma_n \|(\nabla h(x_n) - \nabla h(p_{1,n}), \nabla l_1^*(v_{1,n}) - \nabla l_1^*(p_{2,1,n}), \ldots, \nabla l_m^*(v_{m,n}) - \nabla l_1^*(p_{2,m,n}))\| \\
& = \gamma_n \sqrt{\left\| \sum_{i=1}^m L_i^*(v_{i,n} - p_{2,i,n}) \right\|^2 + \sum_{i=1}^m \|L_i(p_{1,n} - x_n)\|^2} \\
& \quad + \gamma_n \sqrt{\|\nabla h(x_n) - \nabla h(p_{1,n})\|^2 + \sum_{i=1}^m \|\nabla l_i^*(v_{i,n}) - \nabla l_i^*(p_{2,i,n})\|^2} \\
& \leq \gamma_n \sqrt{\left(\sum_{i=1}^m \|L_i\|^2 \right) \sum_{i=1}^m \|v_{i,n} - p_{2,i,n}\|^2 + \left(\sum_{i=1}^m \|L_i\|^2 \right) \|p_{1,n} - x_n\|^2} \\
& \quad + \gamma_n \sqrt{\mu^2 \|x_n - p_{1,n}\|^2 + \sum_{i=1}^m \nu_i^2 \|v_{i,n} - p_{2,i,n}\|^2} \\
& \leq \gamma_n \left(\sqrt{\sum_{i=1}^m \|L_i\|^2} + \max\{\mu, \nu_1, \ldots, \nu_m\} \right) \|(x_n, \boldsymbol{v}_n) - (p_{1,n}, \boldsymbol{p}_{2,n})\|. \qquad (2.15)
\end{aligned}$$

Hence, by taking into consideration the way in which $(\gamma_n)_{n \geq 0}$ is chosen, we have for every $n \geq 0$

$$\begin{aligned}
& \frac{1}{2\gamma_n} \left[\|x_n - p_{1,n}\|^2 + \|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2 - \|x_{n+1} - p_{1,n}\|^2 - \|\boldsymbol{v}_{n+1} - \boldsymbol{p}_{2,n}\|^2 \right] \\
& \geq \frac{1}{2\gamma_n} \left(1 - \gamma_n^2 \left(\sqrt{\sum_{i=1}^m \|L_i\|^2} + \max\{\mu, \nu_1, \ldots, \nu_m\} \right)^2 \right) \|(x, \boldsymbol{v}_n) - (p_{1,n}, \boldsymbol{p}_{2,n})\|^2 \geq 0.
\end{aligned}$$

and, consequently, (2.13) reduces to

$$\frac{\|x_n - x\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \boldsymbol{v}\|^2}{2\gamma_n} \geq \frac{\gamma_{n+1}}{\gamma_n}\frac{\|x_{n+1} - x\|^2}{2\gamma_{n+1}} + [\langle \boldsymbol{L}p_{1,n} - \boldsymbol{r}, \boldsymbol{v}\rangle - G^*(\boldsymbol{v}) + F(p_{1,n})]$$
$$+ \frac{\gamma_{n+1}}{\gamma_n}\frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{v}\|^2}{2\gamma_{n+1}} - \left[\left\langle \boldsymbol{L}x - \boldsymbol{r}, \boldsymbol{p}_{2,n}\right\rangle - G^*(\boldsymbol{p}_{2,n}) + F(x)\right].$$

Let $N \geq 1$ be an arbitrary natural number. Summing the above inequality up from $n = 0$ to $N - 1$ and using the fact that $(\gamma_n)_{n\geq 0}$ is nondecreasing, it follows that

$$\frac{\|x_0 - x\|^2}{2\gamma_0} + \frac{\|\boldsymbol{v}_0 - \boldsymbol{v}\|^2}{2\gamma_0} \geq \frac{\|x_N - x\|^2}{2\gamma_N} + \sum_{n=0}^{N-1}[\langle \boldsymbol{L}p_{1,n} - \boldsymbol{r}, \boldsymbol{v}\rangle - G^*(\boldsymbol{v}) + F(p_{1,n})]$$
$$+ \frac{\|\boldsymbol{v}_N - \boldsymbol{v}\|^2}{2\gamma_N} - \sum_{n=0}^{N-1}\left[\left\langle \boldsymbol{L}x - \boldsymbol{r}, \boldsymbol{p}_{2,n}\right\rangle - G^*(\boldsymbol{p}_{2,n}) + F(x)\right]. \tag{2.16}$$

Replacing $x = \overline{x}$ and $\boldsymbol{v} = \overline{\boldsymbol{v}}$ in the above estimate, since they fulfill (2.5), we obtain

$$\sum_{n=0}^{N-1}[\langle \boldsymbol{L}p_{1,n} - \boldsymbol{r}, \overline{\boldsymbol{v}}\rangle - G^*(\overline{\boldsymbol{v}}) + F(p_{1,n})] - \sum_{n=0}^{N-1}\left[\left\langle \boldsymbol{L}\overline{x} - \boldsymbol{r}, \boldsymbol{p}_{2,n}\right\rangle - G^*(\boldsymbol{p}_{2,n}) + F(\overline{x})\right] \geq 0.$$

Consequently,

$$\frac{\|x_0 - \overline{x}\|^2}{2\gamma_0} + \frac{\|\boldsymbol{v}_0 - \overline{\boldsymbol{v}}\|^2}{2\gamma_0} \geq \frac{\|x_N - \overline{x}\|^2}{2\gamma_N} + \frac{\|\boldsymbol{v}_N - \overline{\boldsymbol{v}}\|^2}{2\gamma_N}$$

and statement *(c)* follows. On the other hand, dividing (2.16) by N, using the convexity of F and G^*, and denoting $x^N := \frac{1}{N}\sum_{n=0}^{N-1} p_{1,n}$ and $v_i^N := \frac{1}{N}\sum_{n=0}^{N-1} p_{2,i,n}$, $i = 1, \dots, m$, we obtain

$$\frac{1}{N}\left(\frac{\|x_0 - x\|^2}{2\gamma_0} + \frac{\|\boldsymbol{v}_0 - \boldsymbol{v}\|^2}{2\gamma_0}\right) \geq \left[\left\langle \boldsymbol{L}x^N - \boldsymbol{r}, \boldsymbol{v}\right\rangle - G^*(\boldsymbol{v}) + F(x^N)\right]$$
$$- \left[\left\langle \boldsymbol{L}x - \boldsymbol{r}, \boldsymbol{v}^N\right\rangle - G^*(\boldsymbol{v}^N) + F(x)\right],$$

which shows (2.8) when passing to the supremum over $x \in B_1$ and $\boldsymbol{v} \in B_2$. In this way statement *(d)* is verified. The weak convergence of $(x^N, \boldsymbol{v}^N)$ to $(\overline{x}, \overline{\boldsymbol{v}})$ when N converges to $+\infty$ is an easy consequence of the Stolz–Cesàro Theorem, fact which shows *(e)*. □

The following remark is formulated in the spirit of [9, Remark 3].

Remark 2.1. In the situation when the functions g_i are Lipschitz continuous on $\mathcal{G}_i, i = 1, ..., m$, inequality (2.8) provides for the sequence of the values of the objective of (P) taken at $(x^N)_{N\geq 1}$ a convergence rate of $\mathcal{O}(\frac{1}{N})$, namely, it holds

$$F(x^N) + G(\boldsymbol{L}x^N - \boldsymbol{r}) - F(\overline{x}) - G(\boldsymbol{L}\overline{x} - \boldsymbol{r}) \leq \frac{C(B_1, B_2)}{N} \ \forall N \geq 1. \tag{2.17}$$

Indeed, due to statement *(b)* of the previous theorem, the sequence $(p_{1,n})_{n\geq 0} \subseteq \mathcal{H}$ is bounded and one can take $B_1 \subset \mathcal{H}$ being a bounded, convex and closed set containing this sequence. Obviously, $\bar{x} \in B_1$. On the other hand, we take $B_2 = \operatorname{dom} g_1^* \times \ldots \times$

$\operatorname{dom} g_m^*$, which is in this situation a bounded set. Then it holds, using the Fenchel-Moreau Theorem and the Young-Fenchel inequality, that

$$\begin{aligned}\mathcal{G}_{B_1\times B_2}(x^N,\boldsymbol{v}^N) &= F(x^N)+G(\boldsymbol{L}x^N-\boldsymbol{r})+G^*(\boldsymbol{v}^N)-\inf_{\tilde{x}\in B_1}\left\{\left\langle \boldsymbol{L}\tilde{x}-\boldsymbol{r},\boldsymbol{v}^N\right\rangle+F(\tilde{x})\right\}\\ &\geq F(x^N)+G(\boldsymbol{L}x^N-\boldsymbol{r})+G^*(\boldsymbol{v}^N)-\left\langle \boldsymbol{L}\overline{x}-\boldsymbol{r},\boldsymbol{v}^N\right\rangle-F(\overline{x})\\ &\geq F(x^N)+G(\boldsymbol{L}x^N-\boldsymbol{r})-F(\overline{x})-G(\boldsymbol{L}\overline{x}-\boldsymbol{r}).\end{aligned}$$

Hence, (2.17) follows by statement *(d)* in Theorem 2.1.

In a similar way, one can show that, whenever f is Lipschitz continuous, (2.8) provides for the sequence of the values of the objective of (D) taken at $(v^N)_{N\geq 1}$ a convergence rate of $\mathcal{O}(\frac{1}{N})$.

Remark 2.2. If $\mathcal{G}_i$, $i=1,\dots,m$, are finite-dimensional real Hilbert spaces, then (2.17) is true, even under the weaker assumption that the convex functions $g_i, i=1,...,m$, have full domain, without necessarily being Lipschitz continuous. The set $B_1\subset\mathcal{H}$ can be chosen as in Remark 2.1, but this time we take $B_2=\bigtimes_{i=1}^m\bigcup_{n\geq 0}\partial g_i\left(L_i p_{1,n}\right)\subset\boldsymbol{\mathcal{G}}$, by noticing also that the functions $g_i, i=1,...,m$, are everywhere subdifferentiable.

The set B_2 is bounded, as for every $i=1,\dots,m$ the set $\bigcup_{n\geq 0}\partial g_i\left(L_i p_{1,n}\right)$ is bounded. Let be $i\in\{1,...,m\}$ fixed. Indeed, as $p_{1,n}\rightharpoonup\overline{x}$, it follows that $L_i p_{1,n}\to L_i\bar{x}$ for $i=1,...,m$. Using the fact that the subdifferential of g_i is a locally bounded operator at $L_i\bar{x}$, the boundedness of $\bigcup_{n\geq 0}\partial g_i\left(L_i p_{1,n}\right)$ follows automatically.

For this choice of the sets B_1 and B_2, by using the same arguments as in the previous remark, it follows that (2.17) is true.

3 Zeros of sums of monotone operators

In this section we turn our attention to the primal-dual monotone inclusion problems formulated in Problem 1.1 with the aim to provide accelerations of the iterative method proposed by Combettes and Pesquet in [10, Theorem 3.1] under the additional strong monotonicity assumptions.

3.1 The case when $A+C$ is strongly monotone

We focus first on the case when $A+C$ is ρ-strongly monotone for some $\rho\in\mathbb{R}_{++}$ and investigate the impact of this assumption on the convergence rate of the sequence of primal iterates. The condition $A+C$ is ρ-strongly monotone is fulfilled when either $A:\mathcal{H}\to 2^{\mathcal{H}}$ or $C:\mathcal{H}\to\mathcal{H}$ is ρ-strongly monotone. In case that A is ρ_1-monotone and C is ρ_2-monotone, the sum $A+C$ is ρ-monotone with $\rho=\rho_1+\rho_2$.

Remark 3.1. The situation when $B_i^{-1}+D_i^{-1}$ is τ_i-strongly monotone with $\tau_i\in\mathbb{R}_{++}$ for $i=1,\dots,m$, which improves the convergence rate of the sequence of dual iterates, can be handled with appropriate modifications.

Due to technical reasons we assume in the following that the operators D_i^{-1} in Problem 1.1 are zero for $i=1,\dots,m$, thus, $D_i(0)=\mathcal{G}_i$ and $D_i(x)=\emptyset$ for $x\neq 0$, for $i=1,...,m$. In Remark 3.2 we show how, by employing the product space approach, the results given in this particular context can be employed when treating the primal-dual pair of monotone inclusions (1.2)-(1.3), however, under the assumption that $D_i^{-1}, i=1,...,m$, are cocoercive. The problem we deal with in this subsection is as follows.

Problem 3.1. Consider the real Hilbert spaces $\mathcal{H}$ and $\mathcal{G}_i, i = 1, ..., m$, $A : \mathcal{H} \to 2^{\mathcal{H}}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ a monotone and μ-Lipschitzian operator for some $\mu \in \mathbb{R}_{++}$. Furthermore, let $z \in \mathcal{H}$ and for every $i \in \{1, \dots, m\}$, let $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be maximally monotone operators and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^* B_i(L_i\overline{x} - r_i) + C\overline{x}, \tag{3.1}$$

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1, \dots, \overline{v}_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \left\{ \begin{array}{l} z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in Ax + Cx \\ \overline{v}_i \in B_i(L_i x - r_i),\ i = 1, \dots, m. \end{array} \right. \tag{3.2}$$

The subsequent algorithm represents an accelerated version of the one given in [10, Theorem 3.1] and relies on the fruitful idea of using a second sequence of variable step length parameters $(\sigma_n)_{n\geq 0} \subseteq \mathbb{R}_{++}$, which, together with the sequence of parameters $(\gamma_n)_{n\geq 0} \subseteq \mathbb{R}_{++}$, play an important role in the convergence analysis. For a similar approach given in the context of a primal-dual forward-backward-type algorithm formulated for primal-dual pairs of convex optimization problems we refer the reader to [9].

Algorithm 3.1. Let $x_0 \in \mathcal{H}$, $(v_{1,0}, \dots, v_{m,0}) \in \boldsymbol{\mathcal{G}}$,

$$\gamma_0 \in \left(0, \min\left\{1, \frac{\sqrt{1+4\rho}}{2(1+2\rho)\mu}\right\}\right) \text{ and } \sigma_0 \in \left(0, \frac{1}{2\gamma_0(1+2\rho)\sum_{i=1}^{m}\|L_i\|^2}\right].$$

Consider the following updates:

$$(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} p_{1,n} = J_{\gamma_n A}\left(x_n - \gamma_n\left(Cx_n + \sum_{i=1}^{m} L_i^* v_{i,n} - z\right)\right) \\ \text{For } i = 1, \dots, m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = J_{\sigma_n B_i^{-1}}\left(v_{i,n} + \sigma_n(L_i x_n - r_i)\right) \\ v_{i,n+1} = \sigma_n L_i(p_{1,n} - x_n) + p_{2,i,n} \end{array} \right. \\ x_{n+1} = \gamma_n \sum_{i=1}^{m} L_i^*(v_{i,n} - p_{2,i,n}) + \gamma_n(Cx_n - Cp_{1,n}) + p_{1,n} \\ \theta_n = 1/\sqrt{1 + 2\rho\gamma_n(1-\gamma_n)},\ \ \gamma_{n+1} = \theta_n\gamma_n,\ \ \sigma_{n+1} = \sigma_n/\theta_n. \end{array} \right. \tag{3.3}$$

Theorem 3.1. *In Problem 3.1 suppose that $A + C$ is ρ-strongly monotone with $\rho \in \mathbb{R}_{++}$ and let $(\overline{x}, \overline{v}_1, \dots, \overline{v}_m) \in \mathcal{H} \times \boldsymbol{\mathcal{G}}$ be a primal-dual solution to Problem 3.1. Then for every $n \geq 0$ it holds*

$$\|x_n - \overline{x}\|^2 + \gamma_n \sum_{i=1}^{m} \frac{\|v_{i,n} - \overline{v}_i\|^2}{\sigma_n} \leq \gamma_n^2 \left(\frac{\|x_0 - \overline{x}\|^2}{\gamma_0^2} + \sum_{i=1}^{m} \frac{\|v_{i,0} - \overline{v}_i\|^2}{\gamma_0\sigma_0} \right), \tag{3.4}$$

where γ_n, $\sigma_n \in \mathbb{R}_{++}$, $x_n \in \mathcal{H}$ and $(v_{1,n}, \dots, v_{m,n}) \in \boldsymbol{\mathcal{G}}$ are the iterates generated by Algorithm 3.1.

Proof. Taking into account the definitions of the resolvents occurring in Algorithm 3.1 we obtain

$$\frac{x_n - p_{1,n}}{\gamma_n} - Cx_n - \sum_{i=1}^{m} L_i^* v_{i,n} + z \in Ap_{1,n}$$

and

$$\frac{v_{i,n} - p_{2,i,n}}{\sigma_n} + L_i x_n - r_i \in B_i^{-1} p_{2,i,n}, i = 1, \dots, m,$$

which, in the light of the updating rules in (3.3), furnishes for every $n \geq 0$

$$\text{and} \quad \begin{aligned} &\frac{x_n - x_{n+1}}{\gamma_n} - \sum_{i=1}^m L_i^* p_{2,i,n} + z \in (A+C)p_{1,n} \\ &\frac{v_{i,n} - v_{i,n+1}}{\sigma_n} + L_i p_{1,n} - r_i \in B_i^{-1} p_{2,i,n}, i = 1, \dots, m. \end{aligned} \tag{3.5}$$

The primal-dual solution $(\overline{x}, \overline{v}_1, \dots, \overline{v}_m) \in \mathcal{H} \times \boldsymbol{\mathcal{G}}$ to Problem 3.1 fulfills (see (1.4), where D_i^{-1} are taken to be zero for $i = 1, ..., m$)

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

Since the sum $A + C$ is ρ-strongly monotone, we have for every $n \geq 0$

$$\left\langle p_{1,n} - \overline{x}, \frac{x_n - x_{n+1}}{\gamma_n} - \sum_{i=1}^m L_i^* p_{2,i,n} + z - \left(z - \sum_{i=1}^m L_i^* \overline{v}_i \right) \right\rangle \geq \rho \|p_{1,n} - \overline{x}\|^2 \tag{3.6}$$

while, due to the monotonicity of $B_i^{-1} : \mathcal{G}_i \to 2^{\mathcal{G}_i}$, we obtain for every $n \geq 0$

$$\left\langle p_{2,i,n} - \overline{v}_i, \frac{v_{i,n} - v_{i,n+1}}{\sigma_n} + L_i p_{1,n} - r_i - (L_i \overline{x} - r_i) \right\rangle \geq 0, \ \ i = 1, \dots, m. \tag{3.7}$$

Further, we set

$$\overline{\boldsymbol{v}} = (\overline{v}_1, \dots, \overline{v}_m), \quad \boldsymbol{v}_n = (v_{1,n}, \dots, v_{m,n}), \quad \boldsymbol{p}_{2,n} = (p_{2,1,n}, \dots, p_{2,m,n}).$$

Summing up the inequalities (3.6) and (3.7), it follows that

$$\begin{aligned} &\left\langle p_{1,n} - \overline{x}, \frac{x_n - x_{n+1}}{\gamma_n} \right\rangle + \left\langle \boldsymbol{p}_{2,n} - \overline{\boldsymbol{v}}, \frac{\boldsymbol{v}_n - \boldsymbol{v}_{n+1}}{\sigma_n} \right\rangle + \left\langle p_{1,n} - \overline{x}, \boldsymbol{L}^*(\overline{\boldsymbol{v}} - \boldsymbol{p}_{2,n}) \right\rangle \\ &+ \left\langle \boldsymbol{p}_{2,n} - \overline{\boldsymbol{v}}, \boldsymbol{L}(p_{1,n} - \overline{x}) \right\rangle \geq \rho \|p_{1,n} - \overline{x}\|^2. \end{aligned} \tag{3.8}$$

and, from here,

$$\left\langle p_{1,n} - \overline{x}, \frac{x_n - x_{n+1}}{\gamma_n} \right\rangle + \left\langle \boldsymbol{p}_{2,n} - \overline{\boldsymbol{v}}, \frac{\boldsymbol{v}_n - \boldsymbol{v}_{n+1}}{\sigma_n} \right\rangle \geq \rho \|p_{1,n} - \overline{x}\|^2 \ \forall n \geq 0. \tag{3.9}$$

In the light of the equations

$$\begin{aligned} \left\langle p_{1,n} - \overline{x}, \frac{x_n - x_{n+1}}{\gamma_n} \right\rangle &= \left\langle p_{1,n} - x_{n+1}, \frac{x_n - x_{n+1}}{\gamma_n} \right\rangle + \left\langle x_{n+1} - \overline{x}, \frac{x_n - x_{n+1}}{\gamma_n} \right\rangle \\ &= \frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} - \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|x_n - \overline{x}\|^2}{2\gamma_n} - \frac{\|x_{n+1} - \overline{x}\|^2}{2\gamma_n}, \end{aligned}$$

and

$$\begin{aligned} \left\langle \boldsymbol{p}_{2,n} - \overline{\boldsymbol{v}}, \frac{\boldsymbol{v}_n - \boldsymbol{v}_{n+1}}{\sigma_n} \right\rangle &= \left\langle \boldsymbol{p}_{2,n} - \boldsymbol{v}_{n+1}, \frac{\boldsymbol{v}_n - \boldsymbol{v}_{n+1}}{\sigma_n} \right\rangle + \left\langle \boldsymbol{v}_{n+1} - \overline{\boldsymbol{v}}, \frac{\boldsymbol{v}_n - \boldsymbol{v}_{n+1}}{\sigma_n} \right\rangle \\ &= \frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n} - \frac{\|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n} + \frac{\|\boldsymbol{v}_n - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} - \frac{\|\boldsymbol{v}_{n+1} - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} \end{aligned}$$

inequality (3.9) reads for every $n \geq 0$

$$\frac{\|x_n - \overline{x}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} \geq \rho\|p_{1,n} - \overline{x}\|^2 + \frac{\|x_{n+1} - \overline{x}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_{n+1} - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} + \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n} - \frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} - \frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n}. \quad (3.10)$$

Using that $2ab \leq \alpha a^2 + \frac{b^2}{\alpha}$ for all $a, b \in \mathbb{R}$, $\alpha \in \mathbb{R}_{++}$, we obtain for $\alpha := \gamma_n$,

$$\begin{aligned} \rho\|p_{1,n} - \overline{x}\|^2 &\geq \rho\left(\|x_{n+1} - \overline{x}\|^2 - 2\|x_{n+1} - \overline{x}\|\|x_{n+1} - p_{1,n}\| + \|x_{n+1} - p_{1,n}\|^2\right) \\ &\geq \frac{2\rho\gamma_n(1-\gamma_n)}{2\gamma_n}\|x_{n+1} - \overline{x}\|^2 - \frac{2\rho(1-\gamma_n)}{2\gamma_n}\|x_{n+1} - p_{1,n}\|^2, \end{aligned}$$

which, in combination with (3.10), yields for every $n \geq 0$

$$\frac{\|x_n - \overline{x}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} \geq \frac{(1 + 2\rho\gamma_n(1-\gamma_n))\|x_{n+1} - \overline{x}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_{n+1} - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} + \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n} - \frac{(1 + 2\rho(1-\gamma_n))\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} - \frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n}. \quad (3.11)$$

Investigating the last two terms in the right-hand side of the above estimate, it shows for every $n \geq 0$ that

$$\begin{aligned} &- \frac{(1 + 2\rho(1-\gamma_n))\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} \\ &\geq -\frac{(1+2\rho)\gamma_n}{2}\left\|\sum_{i=1}^m L_i^*(v_{i,n} - p_{2,i,n}) + (Cx_n - Cp_{1,n})\right\|^2 \\ &\geq -\frac{2(1+2\rho)\gamma_n}{2}\left(\left(\sum_{i=1}^m \|L_i\|^2\right)\|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2 + \mu^2\|x_n - p_{1,n}\|^2\right), \end{aligned}$$

and

$$-\frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n} = -\frac{\sigma_n}{2}\left(\sum_{i=1}^m \|L_i(p_{1,n} - x_n)\|^2\right) \geq -\frac{\sigma_n}{2}\left(\sum_{i=1}^m \|L_i\|^2\right)\|p_{1,n} - x_n\|^2.$$

Hence, for every $n \geq 0$ it holds

$$\begin{aligned} &\frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n} - \frac{(1 + 2\rho(1-\gamma_n))\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} - \frac{\|\boldsymbol{v}_{n+1} - \boldsymbol{p}_{2,n}\|^2}{2\sigma_n} \\ &\geq \frac{(1 - \gamma_n\sigma_n\sum_{i=1}^m \|L_i\|^2 - 2(1+2\rho)\gamma_n^2\mu^2)}{2\gamma_n}\|p_{1,n} - x_n\|^2 \\ &\quad + \frac{(1 - 2\gamma_n\sigma_n(1+2\rho)\sum_{i=1}^m \|L_i\|^2)}{2\sigma_n}\|\boldsymbol{v}_n - \boldsymbol{p}_{2,n}\|^2 \\ &\geq 0. \end{aligned}$$

The nonnegativity of the expression in the above relation follows because of the sequence $(\gamma_n)_{n\geq 0}$ is nonincreasing, $\gamma_n\sigma_n = \gamma_0\sigma_0$ for every $n \geq 0$ and

$$\gamma_0 \in \left(0, \min\left\{1, \frac{\sqrt{1+4\rho}}{2(1+2\rho)\mu}\right\}\right) \text{ and } 0 < \sigma_0 \leq \frac{1}{2\gamma_0(1+2\rho)\sum_{i=1}^m \|L_i\|^2}.$$

Consequently, inequality (3.11) becomes

$$\frac{\|x_n - \overline{x}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_n - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} \geq \frac{(1 + 2\rho\gamma_n(1-\gamma_n))\|x_{n+1} - \overline{x}\|^2}{2\gamma_n} + \frac{\|\boldsymbol{v}_{n+1} - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} \ \forall n \geq 0. \tag{3.12}$$

Dividing (3.12) by γ_n and making use of

$$\theta_n = \frac{1}{\sqrt{1 + 2\rho\gamma_n(1-\gamma_n)}}, \quad \gamma_{n+1} = \theta_n\gamma_n, \quad \sigma_{n+1} = \frac{\sigma_n}{\theta_n},$$

we obtain

$$\frac{\|x_n - \overline{x}\|^2}{2\gamma_n^2} + \frac{\|\boldsymbol{v}_n - \overline{\boldsymbol{v}}\|^2}{2\gamma_n\sigma_n} \geq \frac{\|x_{n+1} - \overline{x}\|^2}{2\gamma_{n+1}^2} + \frac{\|\boldsymbol{v}_{n+1} - \overline{\boldsymbol{v}}\|^2}{2\gamma_{n+1}\sigma_{n+1}} \ \forall n \geq 0.$$

Let be $N \geq 1$. Summing this inequalities from $n = 0$ to $N - 1$, we finally get

$$\frac{\|x_0 - \overline{x}\|^2}{2\gamma_0^2} + \frac{\|\boldsymbol{v}_0 - \overline{\boldsymbol{v}}\|^2}{2\gamma_0\sigma_0} \geq \frac{\|x_N - \overline{x}\|^2}{2\gamma_N^2} + \frac{\|\boldsymbol{v}_N - \overline{\boldsymbol{v}}\|^2}{2\gamma_N\sigma_N}. \tag{3.13}$$

In conclusion,

$$\frac{\|x_n - \overline{x}\|^2}{2} + \gamma_n\frac{\|\boldsymbol{v}_n - \overline{\boldsymbol{v}}\|^2}{2\sigma_n} \leq \gamma_n^2\left(\frac{\|x_0 - \overline{x}\|^2}{2\gamma_0^2} + \frac{\|\boldsymbol{v}_0 - \overline{\boldsymbol{v}}\|^2}{2\gamma_0\sigma_0}\right) \ \forall n \geq 0, \tag{3.14}$$

which completes the proof. □

Next we show that $\rho\gamma_n$ converges like $\frac{1}{n}$ as $n \to +\infty$.

Proposition 3.2. *Let $\gamma_0 \in (0,1)$ and consider the sequence $(\gamma_n)_{n\geq 0} \subseteq \mathbb{R}_{++}$, where*

$$\gamma_{n+1} = \frac{\gamma_n}{\sqrt{1 + 2\rho\gamma_n(1-\gamma_n)}} \ \forall n \geq 0. \tag{3.15}$$

Then $\lim_{n\to+\infty} n\rho\gamma_n = 1$.

Proof. Since the sequence $(\gamma_n)_{n\geq 0} \subseteq (0,1)$ is bounded and decreasing, it converges towards some $l \in [0,1)$ as $n \to +\infty$. We let $n \to +\infty$ in (3.15) and obtain

$$l^2(1 + 2\rho l(1-l)) = l^2 \Leftrightarrow 2\rho l^3(1-l) = 0,$$

which shows that $l = 0$, i.e. $\gamma_n \to 0$ $(n \to +\infty)$. On the other hand, (3.15) implies that $\frac{\gamma_n}{\gamma_{n+1}} \to 1 (n \to +\infty)$. As $(\frac{1}{\gamma_n})_{n\geq 0}$ is a strictly increasing and unbounded sequence, by applying the Stolz–Cesàro Theorem it shows that

$$\begin{aligned}
\lim_{n\to+\infty} n\gamma_n &= \lim_{n\to+\infty} \frac{n}{\frac{1}{\gamma_n}} = \lim_{n\to+\infty} \frac{n+1-n}{\frac{1}{\gamma_{n+1}} - \frac{1}{\gamma_n}} = \lim_{n\to+\infty} \frac{\gamma_n\gamma_{n+1}}{\gamma_n - \gamma_{n+1}} \\
&= \lim_{n\to+\infty} \frac{\gamma_n\gamma_{n+1}(\gamma_n + \gamma_{n+1})}{\gamma_n^2 - \gamma_{n+1}^2} \overset{(3.15)}{=} \lim_{n\to+\infty} \frac{\gamma_n\gamma_{n+1}(\gamma_n + \gamma_{n+1})}{2\rho\gamma_{n+1}^2\gamma_n(1-\gamma_n)} \\
&= \lim_{n\to+\infty} \frac{\gamma_n + \gamma_{n+1}}{2\rho\gamma_{n+1}(1-\gamma_n)} = \lim_{n\to+\infty} \frac{\frac{\gamma_n}{\gamma_{n+1}} + 1}{2\rho(1-\gamma_n)} = \frac{2}{2\rho} = \frac{1}{\rho},
\end{aligned}$$

which completes the proof. □

The following result is a consequence of Theorem 3.1 and Proposition 3.2.

Theorem 3.3. *In Problem 3.1 suppose that $A + C$ is ρ-strongly monotone and let $(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \boldsymbol{\mathcal{G}}$ be a primal-dual solution to Problem 3.1. Then, for any $\varepsilon > 0$, there exists some $n_0 \in \mathbb{N}$ (depending on ε and $\rho\gamma_0$) such that for any $n \geq n_0$*

$$\|x_n - \overline{x}\|^2 \leq \frac{1+\varepsilon}{n^2}\left(\frac{\|x_0 - \overline{x}\|^2}{\rho^2\gamma_0^2} + \sum_{i=1}^m \frac{\|v_{i,0} - \overline{v}_i\|^2}{\rho^2\gamma_0\sigma_0}\right), \tag{3.16}$$

where γ_n, $\sigma_n \in \mathbb{R}_{++}$, $x_n \in \mathcal{H}$ and $(v_{1,n}, \ldots, v_{m,n}) \in \boldsymbol{\mathcal{G}}$ are the iterates generated by Algorithm 3.1.

Remark 3.2. In Algorithm 3.1 and Theorem 3.3 we assumed that $D_i^{-1} = 0$ for $i = 1, \ldots, m$, however, similar statements can be also provided for Problem 1.1 under the additional assumption that the operators $D_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ are such that D_i^{-1} is ν_i^{-1}-cocoercive with $\nu_i \in \mathbb{R}_{++}$ for $i = 1, \ldots, m$. This assumption is in general stronger than assuming that D_i is monotone and D_i^{-1} is ν_i-Lipschitzian for $i = 1, ..., m$. However, it guarantees that D_i is ν_i^{-1}-strongly monotone and maximally monotone for $i = 1, ..., m$ (see [1, Example 20.28, Proposition 20.22 and Example 22.6]). We introduce the Hilbert space $\boldsymbol{\mathcal{H}} = \mathcal{H} \times \boldsymbol{\mathcal{G}}$, the element $\boldsymbol{z} = (z, 0, \ldots, 0) \in \boldsymbol{\mathcal{H}}$ and the maximally monotone operator $\boldsymbol{A} : \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$, $\boldsymbol{A}(x, y_1, \ldots, y_m) = (Ax, D_1 y_1, \ldots, D_m y_m)$ and the monotone and Lipschitzian operator $\boldsymbol{C} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$, $\boldsymbol{C}(x, y_1, \ldots, y_m) = (Cx, 0, \ldots, 0)$. Notice also that $\boldsymbol{A}+\boldsymbol{C}$ is strongly monotone. Furthermore, we introduce the element $\boldsymbol{r} = (r_1, \ldots, r_m) \in \boldsymbol{\mathcal{G}}$, the maximally monotone operator $\boldsymbol{B} : \boldsymbol{\mathcal{G}} \to 2^{\boldsymbol{\mathcal{G}}}$, $\boldsymbol{B}(y_1, \ldots, y_m) = (B_1 y_1, \ldots, B_m y_m)$, and the linear continuous operator $\boldsymbol{L} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{G}}$, $\boldsymbol{L}(x, y_1 \ldots, y_m) = (L_1 x - y_1, \ldots, L_m x - y_m)$, having as adjoint $\boldsymbol{L}^* : \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{H}}$, $\boldsymbol{L}^*(q_1, \ldots, q_m) = (\sum_{i=1}^m L_i^* q_i, -q_1, \ldots, -q_m)$. We consider the primal problem

$$\text{find } \overline{\boldsymbol{x}} = (\overline{x}, \overline{p}_1 \ldots \overline{p}_m) \in \boldsymbol{\mathcal{H}} \text{ such that } \boldsymbol{z} \in \boldsymbol{A}\overline{\boldsymbol{x}} + \boldsymbol{L}^*\boldsymbol{B}\left(\boldsymbol{L}\overline{\boldsymbol{x}} - \boldsymbol{r}\right) + \boldsymbol{C}\overline{\boldsymbol{x}}, \tag{3.17}$$

together with the dual inclusion problem

$$\text{find } \overline{\boldsymbol{v}} \in \boldsymbol{\mathcal{G}} \text{ such that } (\exists \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \left\{ \begin{array}{l} \boldsymbol{z} - \boldsymbol{L}^*\overline{\boldsymbol{v}} \in \boldsymbol{A}\boldsymbol{x} + \boldsymbol{C}\boldsymbol{x} \\ \overline{\boldsymbol{v}} \in \boldsymbol{B}(\boldsymbol{L}\boldsymbol{x} - \boldsymbol{r}) \end{array} \right. . \tag{3.18}$$

We notice that Algorithm 3.1 can be employed for solving this primal-dual pair of monotone inclusion problems and, by separately involving the resolvents of A, B_i and $D_i, i = 1, ..., m$, as for $\gamma \in \mathbb{R}_{++}$

$$J_{\gamma\boldsymbol{A}}(x, y_1, \ldots, y_m) = (J_{\gamma A}x, J_{\gamma D_1}y_1, \ldots, J_{\gamma D_m}y_m) \; \forall (x, y_1, \ldots, y_m) \in \boldsymbol{\mathcal{H}}$$
$$J_{\gamma\boldsymbol{B}}(q_1, \ldots, q_m) = (J_{\gamma B_1}q_1, \ldots, J_{\gamma B_m}q_m) \; \forall (q_1, \ldots, q_m) \in \boldsymbol{\mathcal{G}}.$$

Having $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}) \in \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{G}}$ a primal-dual solution to the primal-dual pair of monotone inclusion problems (3.17)-(3.18), Algorithm 3.1 generates a sequence of primal iterates fulfilling (3.16) in $\boldsymbol{\mathcal{H}}$. Moreover, $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}})$ is a a primal-dual solution to (3.17)-(3.18) if and only if

$$\boldsymbol{z} - \boldsymbol{L}^*\overline{\boldsymbol{v}} \in \boldsymbol{A}\overline{\boldsymbol{x}} + \boldsymbol{C}\overline{\boldsymbol{x}} \text{ and } \overline{\boldsymbol{v}} \in \boldsymbol{B}\left(\boldsymbol{L}\overline{\boldsymbol{x}} - \boldsymbol{r}\right)$$

$$\Leftrightarrow z - \sum_{i=1}^m L_i^*\overline{v}_i \in A\overline{x} + C\overline{x} \text{ and } \overline{v}_i \in D_i\overline{p}_i, \overline{v}_i \in B_i\left(L_i\overline{x} - \overline{p}_i - r_i\right), i = 1, \ldots, m$$

$$\Leftrightarrow z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in A\overline{x} + C\overline{x} \text{ and } \overline{v}_i \in D_i \overline{p}_i, L_i\overline{x} - r_i \in B_i^{-1}\overline{v}_i + \overline{p}_i, i = 1, \ldots, m.$$

Thus, if $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}})$ is a primal-dual solution to (3.17)-(3.18), then $(\overline{x}, \overline{\boldsymbol{v}})$ is a primal-dual solution to Problem 1.1. Viceversa, if $(\overline{x}, \overline{\boldsymbol{v}})$ is a primal-dual solution to Problem 1.1, then, choosing $\overline{p}_i \in D_i^{-1}\overline{v}_i, i = 1, ..., m$, and $\overline{\boldsymbol{x}} = (\overline{x}, \overline{p}_1 \ldots \overline{p}_m)$, it yields that $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}})$ is a primal-dual solution to (3.17)-(3.18). In conclusion, the first component of every primal iterate in $\boldsymbol{\mathcal{H}}$ generated by Algorithm 3.1 for finding the primal-dual solution $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}})$ to (3.17)-(3.18) will furnish a sequence of iterates verifying (3.16) in $\mathcal{H}$ for the primal-dual solution $(\overline{x}, \overline{\boldsymbol{v}})$ to Problem 1.1.

3.2 The case when $A + C$ and $B_i^{-1} + D_i^{-1}$, $i = 1, \ldots, m$, are strongly monotone

Within this subsection we consider the case when $A + C$ is ρ-strongly monotone with $\rho \in \mathbb{R}_{++}$ and $B_i^{-1} + D_i^{-1}$ is τ_i-strongly monotone with $\tau_i \in \mathbb{R}_{++}$ for $i = 1, \ldots, m$, and provide an accelerated version of the algorithm in [10, Theorem 3.1] which generates sequences of primal and dual iterates that converge to the primal-dual solution to Problem 1.1 with an improved rate of convergence. The provided algorithm and its convergence properties are formulated in the spirit to the investigations in [9, Subsection 5.2].

Algorithm 3.2. Let $x_0 \in \mathcal{H}$, $(v_{1,0}, \ldots, v_{m,0}) \in \boldsymbol{\mathcal{G}}$, and $\gamma \in (0, 1)$ such that

$$\gamma \leq \frac{1}{\sqrt{1 + 2\min\{\rho, \tau_1, \ldots, \tau_m\}}\left(\sqrt{\sum_{i=1}^m \|L_i\|^2} + \max\{\mu, \nu_1, \ldots, \nu_m\}\right)}.$$

Consider the following updates:

$$(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} p_{1,n} = J_{\gamma A}\left(x_n - \gamma\left(Cx_n + \sum_{i=1}^m L_i^* v_{i,n} - z\right)\right) \\ \text{For } i = 1, \ldots, m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = J_{\gamma B_i^{-1}}\left(v_{i,n} + \gamma(L_i x_n - D_i^{-1} v_{i,n} - r_i)\right) \\ v_{i,n+1} = \gamma L_i(p_{1,n} - x_n) + \gamma(D_i^{-1}v_{i,n} - D_i^{-1}p_{2,i,n}) + p_{2,i,n} \end{array} \right. \\ x_{n+1} = \gamma \sum_{i=1}^m L_i^*(v_{i,n} - p_{2,i,n}) + \gamma(Cx_n - Cp_{1,n}) + p_{1,n}. \end{array} \right. \tag{3.19}$$

Theorem 3.4. *In Problem 1.1 suppose that $A + C$ is ρ-strongly monotone with $\rho \in \mathbb{R}_{++}$, $B_i^{-1} + D_i^{-1}$ is τ_i-strongly monotone with $\tau_i \in \mathbb{R}_{++}$ for $i = 1, \ldots, m$, and let $(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \boldsymbol{\mathcal{G}}$ be a primal-dual solution to Problem 1.1. Then for every $n \geq 0$ it holds*

$$\|x_n - \overline{x}\|^2 + \sum_{i=1}^m \|v_{i,n} - \overline{v}_i\|^2 \leq \left(\frac{1}{1 + 2\rho_{\min}\gamma(1-\gamma)}\right)^n \left(\|x_0 - \overline{x}\|^2 + \sum_{i=1}^m \|v_{i,0} - \overline{v}_i\|^2\right),$$

where $\rho_{\min} = \min\{\rho, \tau_1, \ldots, \tau_m\}$ and $x_n \in \mathcal{H}$ and $(v_{1,n}, \ldots, v_{m,n}) \in \boldsymbol{\mathcal{G}}$ are the iterates generated by Algorithm 3.2.

Proof. Taking into account the definitions of the resolvents occurring in Algorithm 3.2 and the fact that the primal-dual solution $(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \boldsymbol{\mathcal{G}}$ to Problem 1.1 fulfills (1.4), by the strong monotonicity of $A + C$ and $B_i^{-1} + D_i^{-1}$, $i = 1, \ldots, m$, we obtain for every $n \geq 0$

$$\left\langle p_{1,n} - \overline{x}, \frac{x_n - x_{n+1}}{\gamma} - \sum_{i=1}^m L_i^* p_{2,i,n} + z - \left(z - \sum_{i=1}^m L_i^* \overline{v}_i\right)\right\rangle \geq \rho\|p_{1,n} - \overline{x}\|^2 \tag{3.20}$$

and, respectively,

$$\left\langle p_{2,i,n} - \overline{v}_i, \frac{v_{i,n} - v_{i,n+1}}{\gamma} + L_i p_{1,n} - r_i - (L_i \overline{x} - r_i) \right\rangle \geq \tau_i \|p_{2,i,n} - \overline{v}_i\|^2, i = 1, ..., m. \tag{3.21}$$

Consider the Hilbert space $\boldsymbol{\mathcal{H}} = \mathcal{H} \times \boldsymbol{\mathcal{G}}$, equipped with the inner product defined in (2.14) and associated norm, and set

$$\overline{\boldsymbol{x}} = (\overline{x}, \overline{v}_1, \ldots, \overline{v}_m), \quad \boldsymbol{x}_n = (x_n, v_{1,n}, \ldots, v_{m,n}), \quad \boldsymbol{p}_n = (p_{1,n}, p_{2,1,n}, \ldots, p_{2,m,n}).$$

Summing up the inequalities (3.20) and (3.21) and using

$$\left\langle \boldsymbol{p}_n - \overline{\boldsymbol{x}}, \frac{\boldsymbol{x}_n - \boldsymbol{x}_{n+1}}{\gamma} \right\rangle = \frac{\|\boldsymbol{x}_{n+1} - \boldsymbol{p}_n\|^2}{2\gamma} - \frac{\|\boldsymbol{x}_n - \boldsymbol{p}_n\|^2}{2\gamma} + \frac{\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}\|^2}{2\gamma} - \frac{\|\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}\|^2}{2\gamma},$$

we obtain for every $n \geq 0$

$$\frac{\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}\|^2}{2\gamma} \geq \rho_{\min} \|\boldsymbol{p}_n - \overline{\boldsymbol{x}}\|^2 + \frac{\|\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}\|^2}{2\gamma} + \frac{\|\boldsymbol{x}_n - \boldsymbol{p}_n\|^2}{2\gamma} - \frac{\|\boldsymbol{x}_{n+1} - \boldsymbol{p}_n\|^2}{2\gamma}. \tag{3.22}$$

Further, using the estimate $2ab \leq \gamma a^2 + \frac{b^2}{\gamma}$ for all $a, b \in \mathbb{R}$, we obtain

$$\begin{aligned} \rho_{\min} \|\boldsymbol{p}_n - \overline{\boldsymbol{x}}\|^2 &\geq \frac{2\rho_{\min}\gamma(1-\gamma)}{2\gamma} \|\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}\|^2 - \frac{2\rho_{\min}(1-\gamma)}{2\gamma} \|\boldsymbol{x}_{n+1} - \boldsymbol{p}_n\|^2 \\ &\geq \frac{2\rho_{\min}\gamma(1-\gamma)}{2\gamma} \|\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}\|^2 - \frac{2\rho_{\min}}{2\gamma} \|\boldsymbol{x}_{n+1} - \boldsymbol{p}_n\|^2 \ \forall n \geq 0. \end{aligned}$$

Hence, (3.22) reduces to

$$\begin{aligned} \frac{\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}\|^2}{2\gamma} \geq & \frac{(1 + 2\rho_{\min}\gamma(1-\gamma))\|\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}\|^2}{2\gamma} \\ & + \frac{\|\boldsymbol{x}_n - \boldsymbol{p}_n\|^2}{2\gamma} - \frac{(1 + 2\rho_{\min})\|\boldsymbol{x}_{n+1} - \boldsymbol{p}_n\|^2}{2\gamma} \ \forall n \geq 0. \end{aligned}$$

Using the same arguments as in (2.15), it is easy to check that for every $n \geq 0$

$$\begin{aligned} & \frac{\|\boldsymbol{x}_n - \boldsymbol{p}_n\|^2}{2\gamma} - \frac{(1 + 2\rho_{\min})\|\boldsymbol{x}_{n+1} - \boldsymbol{p}_n\|^2}{2\gamma} \\ & \geq \left(1 - (1 + 2\rho_{\min})\gamma^2 \left(\sqrt{\sum_{i=1}^m \|L_i\|^2} + \max\{\mu, \nu_1, \ldots, \nu_m\}\right)^2\right) \frac{\|\boldsymbol{x}_n - \boldsymbol{p}_n\|^2}{2\gamma} \\ & \geq 0, \end{aligned}$$

whereby the nonnegativity of this term is ensured by the assumption that

$$\gamma \leq \frac{1}{\sqrt{1 + 2\rho_{\min}}\left(\sqrt{\sum_{i=1}^m \|L_i\|^2} + \max\{\mu, \nu_1, \ldots, \nu_m\}\right)}.$$

Therefore, we obtain

$$\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}\|^2 \geq (1 + 2\rho_{\min}\gamma(1-\gamma))\|\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}\|^2 \ \forall n \geq 0,$$

which leads to

$$\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}\|^2 \leq \left(\frac{1}{1 + 2\rho_{\min}\gamma(1-\gamma)}\right)^n \|\boldsymbol{x}_0 - \overline{\boldsymbol{x}}\|^2 \ \forall n \geq 0.$$

□

4 Numerical experiments in imaging

In this section we test the feasibility of Algorithm 2.1 and of its accelerated version Algorithm 3.1 in the context of different problem formulations occurring in imaging and compare their performances to the ones of several popular algorithms in the field. For all applications discussed in this section the images have been normalized, in order to make their pixels range in the closed interval from 0 to 1.

4.1 TV-based image denoising

Our first numerical experiment aims the solving of an image denoising problem via total variation regularization. More precisely, we deal with the convex optimization problem

$$\inf_{x\in\mathbb{R}^n}\left\{\lambda TV(x)+\frac{1}{2}\|x-b\|^2\right\}, \tag{4.1}$$

where $\lambda\in\mathbb{R}_{++}$ is the regularization parameter, $TV:\mathbb{R}^n\to\mathbb{R}$ is a discrete total variation functional and $b\in\mathbb{R}^n$ is the observed noisy image.

In this context, $x\in\mathbb{R}^n$ represents the vectorized image $X\in\mathbb{R}^{M\times N}$, where $n=M\cdot N$ and $x_{i,j}$ denotes the normalized value of the pixel located in the i-th row and the j-th column, for $i=1,\dots,M$ and $j=1,\dots,N$. We denote $\mathcal{Y}=\mathbb{R}^n\times\mathbb{R}^n$ and define the linear operator $L:\mathbb{R}^n\to\mathcal{Y}$, $x_{i,j}\mapsto(L_1x_{i,j},L_2x_{i,j})$, where

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

The operator L represents a discretization of the gradient using reflexive (Neumann) boundary conditions and standard finite differences and fulfills $\|L\|^2\leq 8$. Its adjoint $L^*:\mathcal{Y}\to\mathbb{R}^n$ is given in [8].

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

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

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

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

where in both cases reflexive (Neumann) boundary conditions are assumed.

Within this example we will focus on the anisotropic total variation function which is nothing else than the composition of the l_1-norm in $\mathcal{Y}$ with the linear operator L. Due to the full splitting characteristics of the iterative methods presented in this paper, we need only to compute the proximal point of the conjugate of the l_1-norm, the latter being the

	$\sigma = 0.12,\ \lambda = 0.07$		$\sigma = 0.06,\ \lambda = 0.035$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$
ALG1	7.51s (343)	49.66s (2271)	4.08s (187)	34.44s (1586)
ALG2	2.20s (101)	9.84s (451)	1.61s (73)	6.70s (308)
PD1	3.69s (337)	24.34s (2226)	2.02s (183)	16.74s (1532)
PD2	1.08s (96)	4.94s (447)	0.79s (70)	3.53s (319)
AMA	5.07s (471)	32.59s (3031)	2.74s (254)	23.49s (2184)
Fast AMA	1.06s (89)	6.63s (561)	0.75s (63)	4.53s (383)
Nesterov	1.15s (102)	6.66s (595)	0.81s (72)	4.70s (415)
FISTA	0.96s (100)	6.12s (645)	0.68s (70)	4.08s (429)

Table 4.1: Performance evaluation for the images in Figure 4.1. The entries represent 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.

indicator function of the dual unit ball. Thus, the calculation of the proximal point will result in the computation of a projection, which has an easy implementation. The more challenging isotropic total variation functional is employed in the forthcoming subsection in the context of an image deblurring problem.

Thus, problem (4.1) reads equivalently

$$\inf_{x \in \mathbb{R}^n} \left\{ h(x) + g(Lx) \right\},$$

where $h : \mathbb{R}^n \to \mathbb{R}$, $h(x) = \frac{1}{2}\|x - b\|^2$, is 1-strongly monotone and differentiable with 1-Lipschitzian gradient and $g : \mathcal{Y} \to \mathbb{R}$ is defined as $g(y_1, y_2) = \lambda\|(y_1, y_2)\|_1$. Then its conjugate $g^* : \mathcal{Y} \to \overline{\mathbb{R}}$ is nothing else than

$$g^*(p_1, p_2) = (\lambda\| \cdot \|_1)^* (p_1, p_2) = \lambda \left\| \left(\frac{p_1}{\lambda}, \frac{p_2}{\lambda} \right) \right\|_1^* = \delta_S(p_1, p_2),$$

where $S = [-\lambda, \lambda]^n \times [-\lambda, \lambda]^n$.

We solved the regularized image denoising problem with Algorithm 2.1 and Algorithm 3.1 and other first-order methods. A comparison of the obtained results is shown in Table 4.1, where the abbreviations refer to the following algorithms (for which the initial parameter choices are also specified):

- ALG1: Algorithm 2.1 with $\gamma = \frac{1-\tilde{\varepsilon}}{\sqrt{8}}$, small $\tilde{\varepsilon} > 0$ and by taking the last iterate instead of the averaged sequence.
- ALG2: Algorithm 3.1 with $\rho = 0.3$, $\mu = 1$ and $\gamma_0 = \frac{\sqrt{1+4\rho}}{2(1+2\rho)\mu}$.
- PD1: Algorithm 1 in [9] with $\tau = \frac{1}{\sqrt{8}}$, $\tau\sigma 8 = 1$ and by taking the last iterate instead of the averaged sequence.
- PD2: Algorithm 2 in [9] with $\rho = 0.3$, $\tau_0 = \frac{1}{\sqrt{8}}$, $\tau_0\sigma_0 8 = 1$.
- AMA: The scheme (4.3a)–(4.3c) in [14] for $c(t) = \frac{2}{\|L\|^2}$ for all $t \geq 1$.
- Fast AMA: The AMA scheme for $c(t) = \frac{1.2}{\|L\|^2}$ for all $t \geq 1$ and FISTA-type acceleration.
- Nesterov: The scheme (3.11) in [12] on the dual problem.
- FISTA: The scheme (4.1)–(4.3) in [2] on the dual problem.

Figure 4.1: TV-l_2 image denoising. The noisy image in (a) was obtained after adding white Gaussian noise with standard deviation $\sigma = 0.06$ to the original 256×256 lichtenstein test image, while the output of Algorithm 3.1, for $\lambda = 0.035$, after 100 iterations is shown in (c). Likewise, the noise image when choosing $\sigma = 0.12$ and the output of the same algorithm, for $\lambda = 0.07$, after 100 iterations are shown in (b) and (d), respectively.

From the point of view of the number of iterations, one can notice similarities between both the primal-dual algorithms ALG1 and PD1 and the accelerated versions ALG2 and PD2. With regard to this criterion they behave almost equal. When comparing the CPU times, it shows that the methods in this paper need almost twice amount of time. This is since ALG1 and ALG2 lead back to a forward-backward-forward splitting scheme, whereas PD1 and PD2 rely on a forward-backward splitting scheme, meaning that ALG1 and ALG2 process the double amount of forward steps than PD1 and PD2 (and, actually, any other algorithm listed in Table 4.1). In the considered numerical experiment the evaluation of the forward steps (which are actually matix-vector multiplications involving the linear operators and their adjoints) is, compared with the calculation of projections when computing the resolvents, the most costly step. However, in contrast to all other algorithms listed in Table 4.1, ALG1 and ALG2 are parallelizable. Hence, in distributed programming, two matrix-vector products can be calculated simultaneously, which reduces the time per iteration to a level where the other algorithms already are. In this situation, the number of iterations becomes a more important feature, and this is an aspect where our accelerated method ALG2 shows to be competitive to state-of-the-art solvers. It is also noticable, that the acceleration of the alternating minimization algorithm (AMA) as well as the accelerated first-order methods operating on the dual problem, i. e. Nesterov and FISTA, perform very well on this example.

4.2 TV-based image deblurring

The second numerical experiment that we consider concerns solving an extremely ill-conditioned linear inverse problem which arises in image deblurring. For a given matrix

$A \in \mathbb{R}^{n\times n}$ describing a blur operator and a given vector $b \in \mathbb{R}^n$ representing the blurred and noisy image, the task is to estimate the unknown original image $\overline{x} \in \mathbb{R}^n$ fulfilling

$$A\overline{x} = b.$$

To this end we basically solve the following regularized convex nondifferentiable problem

$$\inf_{x\in\mathbb{R}^n} \left\{ \|Ax - b\|_1 + \lambda TV_{\text{iso}}(x) + \delta_{[0,1]^n}(x) \right\}, \tag{4.2}$$

where $\lambda \in \mathbb{R}_{++}$ is a regularization parameter and $TV_{\text{iso}} : \mathbb{R}^n \to \mathbb{R}$ is the discrete isotropic total variation function. Notice that none of the functions occurring in (4.2) is differentiable.

For (y, z), $(p, q) \in \mathcal{Y}$, we introduce the inner product

$$\langle (y, z), (p, q)\rangle = \sum_{i=1}^{M}\sum_{j=1}^{N} y_{i,j}p_{i,j} + z_{i,j}q_{i,j}$$

and define $\|(y, z)\|_\times = \sum_{i=1}^{M}\sum_{j=1}^{N} \sqrt{y_{i,j}^2 + z_{i,j}^2}$. One can check that $\|\cdot\|_\times$ is a norm on $\mathcal{Y}$ and that for every $x \in \mathbb{R}^n$ it holds $TV_{\text{iso}}(x) = \|Lx\|_\times$, where L is the linear operator defined in the previous section. The conjugate function $(\|\cdot\|_\times)^* : \mathcal{Y} \to \overline{\mathbb{R}}$ of $\|\cdot\|_\times$ is for every $(p, q) \in \mathcal{Y}$ given by

$$(\|\cdot\|_\times)^*(p, q) = \begin{cases} 0, & \text{if } \|(p, q)\|_{\times*} \le 1 \\ +\infty, & \text{otherwise} \end{cases},$$

where

$$\|(p, q)\|_{\times*} = \sup_{\|(y,z)\|_\times \le 1} \langle (p, q), (y, z)\rangle = \max_{\substack{1\le i\le M\\ 1\le j\le N}} \sqrt{p_{i,j}^2 + q_{i,j}^2}.$$

Therefore, the optimization problem (4.2) can be written in the form of

$$\inf_{x\in\mathbb{R}^n} \{f(x) + g_1(Ax) + g_2(Lx)\},$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $f(x) = \delta_{[0,1]^n}(x)$, $g_1 : \mathbb{R}^n \to \mathbb{R}$, $g_1(y) = \|y - b\|_1$, and $g_2 : \mathcal{Y} \to \mathbb{R}$, $g_2(y, z) = \lambda \|(y, z)\|_\times$. For every $p \in \mathbb{R}^n$, it holds $g_1^*(p) = \delta_{[-1,1]^n}(p) + p^T b$ (see, for instance, [3]), while for any $(p, q) \in \mathcal{Y}$, we have $g_2^*(p, q) = \delta_S(p, q)$, with $S = \{(p, q) \in \mathcal{Y} : \|(p, q)\|_{\times*} \le \lambda\}$. We solved this problem by Algorithm 2.1 and to this end we made use of the following formulae

$$\begin{array}{l} \text{Prox}_{\gamma f}(x) = \mathcal{P}_{[0,1]^n}(x) \;\; \forall x \in \mathbb{R}^n \\ \text{Prox}_{\gamma g_1^*}(p) = \mathcal{P}_{[-1,1]^n}(p - \gamma b) \;\; \forall p \in \mathbb{R}^n \end{array}, \text{ and } \text{Prox}_{\gamma g_2^*}(p, q) = \mathcal{P}_S(p, q) \;\; \forall (p, q) \in \mathcal{Y},$$

where $\gamma \in \mathbb{R}_{++}$ and the projection operator $\mathcal{P}_S : \mathcal{Y} \to S$ is defined as

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

Figure 4.2 shows the cameraman test image obtained after multiplying the original one with the blur operator and adding normally distributed white Gaussian noise with standard deviation 10^{-3}. It also shows the image reconstructed by Algorithm 2.1 when taking as regularization parameter $\lambda = 0.003$. Finally, plots on the function values give an insight into the benefits of taking averaged iterates into consideration which, due to Theorem 2.1, achieve a rate of convergence of $\mathcal{O}(\frac{1}{n})$.

(a) Blurred and noisy image (b) Reconstructed image (c) Relative function values

Figure 4.2: TV-l_1 image deblurring. Figure (a) shows the blurred and noisy 256×256 cameraman test image, (b) shows the averaged iterate generated by Algorithm 2.1 after 400 iterations and (c) shows the relative error in terms of function values when taking the last or the averaged iterate.

4.3 TV-based image inpainting

In the last section of the paper we show how image inpainting problems, which aim for recovering lost information, can be efficiently solved via the primal-dual methods investigated in this work. To this end, we consider the following TV-regularized model

$$\begin{aligned} \inf \quad & TV_{\text{iso}}(x), \\ \text{s.t.} \quad & Kx = b \\ & x \in [0,1]^n \end{aligned} \tag{4.3}$$

where $TV_{\text{iso}} : \mathbb{R}^n \to \mathbb{R}$ is the isotropic total variation functional and $K \in \mathbb{R}^{n \times n}$ is the diagonal matrix, where for $i = 1, ..., n$, $K_{i,i} = 0$, if the pixel i in the noisy image $b \in \mathbb{R}^n$ is lost (in our case set to black) and $K_{i,i} = 1$, otherwise. The induced linear operator $K : \mathbb{R}^n \to \mathbb{R}^n$ fulfills $\|K\| = 1$, while, in the light of the considerations made in the previous two subsections, we have that $TV_{\text{iso}}(x) = \|Lx\|_\times$ for all $x \in \mathbb{R}^n$.

Thus, problem (4.3) can be formulated as

$$\inf_{x \in \mathbb{R}^n} \{f(x) + g_1(Lx) + g_2(Kx)\},$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $f(x) = \delta_{[0,1]^n}$, $g_1 : \mathcal{Y} \to \mathbb{R}$, $g_1(y_1, y_2) = \|(y_1, y_2)\|_\times$ and $g_2 : \mathbb{R}^n \to \mathbb{R}$, $g_2(y) = \delta_{\{0\}}(y - b)$. We solve it by Algorithm 2.1, the formulae for the proximal points involved in this iterative scheme been already given in Subsection 4.2. Figure 4.3 shows the original fruit image, the image obtained from it after setting 80% randomly chosen pixels to black and the image reconstructed by Algorithm 2.1.

Acknowledgements. The authors are thankful to the anonymous reviewers for pertinent comments and suggestions which improved the quality of the paper.

References

[1] H.H. Bauschke and 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.

(a) Original image (b) 80% missing pixels (c) Reconstructed image

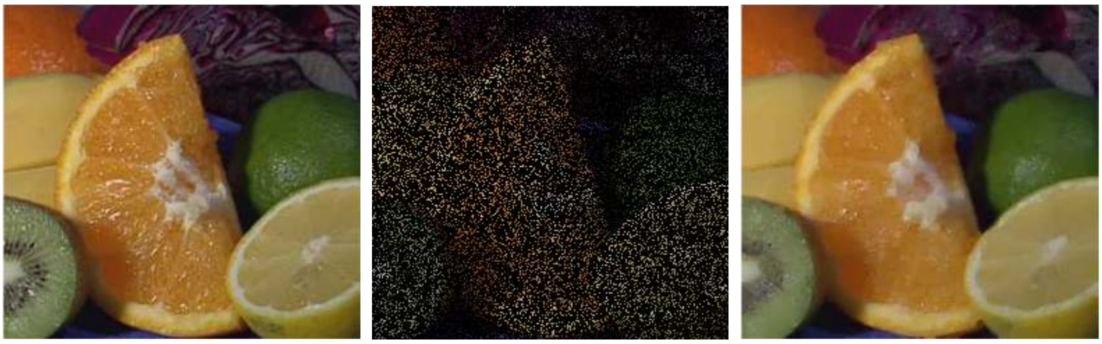

Figure 4.3: *TV* image inpainting. Figure (a) shows the 240×256 clean fruits image, (b) shows the same image for which 80% randomly chosen pixels were set to black and (c) shows the nonaveraged iterate generated by Algorithm 2.1 after 200 iterations.

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

[4] R.I. Boţ, E.R. Csetnek and A. Heinrich. A primal-dual splitting algorithm for finding zeros of sums of maximally monotone operators. *SIAM J. Optim.* **23**(4), 2011–2036, 2013.

[5] R.I. Boţ, E.R. Csetnek and A. Heinrich. On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems. *arXiv:1303.2875*, 2013.

[6] 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. *SIAM J. Optim.*, **23**(4), 2541–2565, 2013.

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

[8] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**(1–2), 89–97, 2004.

[9] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145, 2011.

[10] P.L. Combettes and J.-C. Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.* **20**(2), 307–330, 2012.

[11] L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.* **158**(2), 460–479, 2013.

[12] Y. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.* **103**(1), 127–152, 2005.

[13] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.* **38**(2), 431–446, 2000.

[14] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Optim.* **29**(1), 119–138, 1991.

[15] B.C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comp. Math.* **38**(3), 667–681, 2013.

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

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