Recent developments on primal-dual splitting methods with applications to convex minimization

Radu Ioan Boţ, Ernö Robert Csetnek and Christopher Hendrich

Abstract This article presents a survey on primal-dual splitting methods for solving monotone inclusion problems involving maximally monotone operators, linear compositions of parallel sums of maximally monotone operators and single-valued Lipschitzian or cocoercive monotone operators. The primal-dual algorithms 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 or by backward steps for the set-valued ones, by using the corresponding resolvents. In the hypothesis that strong monotonicity assumptions for some of the involved operators are fulfilled, accelerated algorithmic schemes are presented and analyzed from the point of view of their convergence. Finally, we discuss the employment of the primal-dual methods in the context of solving convex optimization problems arising in the fields of image denoising and deblurring, support vector machine learning, location theory, portfolio optimization and clustering.

Key words: maximally monotone operator, resolvent, operator splitting, convergence analysis, convex optimization, subdifferential, numerical experiments

---

Radu Ioan Boţ
University of Vienna, Faculty of Mathematics, Nordbergstraße 15, A-1090 Vienna, Austria, e-mail: radu.bot@univie.ac.at. Research partially supported by DFG (German Research Foundation), project BO 2516/4-1.

Ernö Robert Csetnek
University of Vienna, Faculty of Mathematics, Nordbergstraße 15, A-1090 Vienna, Austria, e-mail: ernoe.robert.csetnek@univie.ac.at. Research supported by DFG (German Research Foundation), project BO 2516/4-1.

Christopher Hendrich
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.

1 Introduction

In the last couple of years a particular attention was given to the development of a new class of so-called primal-dual splitting methods for solving monotone inclusion problems, especially, when they involve mixtures of linearly composed maximally monotone operators, parallel sums of maximally monotone operators and/or single-valued Lipschitzian or cocoercive monotone operators. The efforts done in this sense were motivated by the fact that a wide variety of convex optimization problems such as location problems, support vector machine problems for classication and regression, problems in clustering and portfolio optimization as well as signal and image processing problems, all of them potentially possessing nonsmooth terms in their objectives, can be reduced to the solving of monotone inclusion problems with such an intricate formulation. The classical splitting algorithms, like the forward-backward algorithm [2], Tseng's forward-backward-forward algorithm [39] and the Douglas-Rachford algorithm [2, 25] have considerable limitations when employed on monotone inclusion problems with such an intricate formulation, as they would assume the calculation of the resolvents of linearly composed maximally monotone operators or of parallel sums of maximally monotone operators, for which exact formulae are available only in very exceptional situations (see [2]).

In order to overcome this shortcoming, the primal-dual splitting algorithms solve actually the primal-dual pair formed by the monotone inclusion problem under investigation and its dual inclusion problem in the sense of Attouch-Théra ([1, 2]) by reformulating it as a monotone inclusion problem in a corresponding product space. The algorithmic scheme follows by applying in an appropriate way one of the standard splitting algorithms and 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.

After presenting in the next section some elements of convex analysis and of the theory of maximally monotone operators, we present in Section 3 three main classes of primal-dual splitting algorithms for solving monotone inclusion problems having an intricate formulation along with corresponding convergence statements and discuss possible accelerations, provided that some of the involved operators fulfill strong monotonicity assumptions.

In the hypothesis that the single-valued monotone operator arising in the formulation of the monotone inclusion problem is cocoercive, we present first an adaptation of the primal-dual algorithm proposed by Vũ in [40], that relies on the employment of the forward-backward splitting method in an appropriate product space. For particular instances of this iterative scheme in the context of monotone inclusion problems we refer the reader to [8] and in the context of convex optimization problems to [19, 24]. Further, we discuss two accelerated versions of it proposed in [9], for which an evaluation of the convergence behaviour of the sequences of primal and dual iterates, respectively, is possible.

In Subsection 3.2, provided that the single-valued monotone operator arising in the formulation of the monotone inclusion problem is Lipschitzian, we turn out our attention to a primal-dual method due to Combettes and Pesquet ([23]; see, also, [17]), which can be reduced to Tseng's forward-backward-forward splitting method in a product space. Two accelerated versions of the forward-backward-forward type primal-dual algorithm introduced in [13] are presented under strong monotonicity assumptions, as well, along with the corresponding convergence statements.

In the last subsection of Section 3 we present two primal-dual methods proposed in [14] that rely on the Douglas-Rachford splitting algorithm in a product space and discuss their convergence behaviour.

In the last part of the article we discuss the employment of the presented primal-dual methods in the context of solving convex optimization problems. Numerical experiments are made in the context of applications arising in the fields of image denoising and deblurring, support vector machine learning, location theory, portfolio optimization and clustering.

2 Preliminaries

Let us start by presenting some notations which are used throughout the work (see [2, 6, 7, 26, 37, 41]). We consider real Hilbert spaces $\mathcal{H}$ and $\mathcal{G}_i, i = 1,...,m$, 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, $\mathbb{R}_{++}$ denotes the set of strictly positive real numbers and $\mathbb{R}_+ = \mathbb{R}_{++} \cup \{0\}$. By $B(0,r)$ we denote the closed ball with center 0 and radius $r \in \mathbb{R}_{++}$. For a function $f : \mathcal{H} \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ 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 $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ 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 $L_i : \mathcal{H} \rightarrow \mathcal{G}_i$, the operator $L_i^* : \mathcal{G}_i \rightarrow \mathcal{H}$, 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*, for $i = 1,\ldots,m$. Having two proper functions $f, g : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, their *infimal convolution* is defined by $f \Box g : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, $(f \Box g)(x) = \inf_{y \in \mathcal{H}} \{f(y) + g(x-y)\}$ for all $x \in \mathcal{H}$.

Let $M : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a set-valued operator. We denote by $\operatorname{zer} M = \{x \in \mathcal{H} : 0 \in Mx\}$ its set of *zeros*, by $\operatorname{fix} M = \{x \in \mathcal{H} : x \in Mx\}$ its set of *fixed points*, 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* of M is $M^{-1} : \mathcal{H} \rightarrow 2^{\mathcal{H}}$, $u \mapsto \{x \in \mathcal{H} : u \in Mx\}$. We say that the operator M is *monotone*, if $\langle x-y, u-v \rangle \geq 0$ for all $(x,u), (y,v) \in \operatorname{gra} M$ and it is said to be *maximally monotone*, if there exists no monotone operator $M' : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ such that $\operatorname{gra} M'$ properly contains $\operatorname{gra} M$. The operator M is said to be *uniformly monotone* with modulus $\phi_M : \mathbb{R}_+ \to [0,+\infty]$, if ϕ_M is increasing, vanishes only at 0, and $\langle x-y, u-v\rangle \geq \phi_M(\|x-y\|)$ for all $(x,u), (y,v) \in \operatorname{gra} M$. A prominent representative of the class of uniformly monotone operators are the *strongly monotone* ones. Let $\gamma > 0$ be arbitrary. We say that M is γ-strongly monotone, if $\langle x-y, u-v\rangle \geq \gamma\|x-y\|^2$ for all $(x,u), (x,v) \in \operatorname{gra} M$. A single-valued operator $M : \mathcal{H} \to \mathcal{H}$ is said to be *γ-cocoercive*, if $\langle x-y, Mx-My\rangle \geq \gamma\|Mx-My\|^2$ for all $(x,y) \in \mathcal{H} \times \mathcal{H}$. Moreover, M is *γ-Lipschitzian*, if $\|Mx-My\| \leq \gamma\|x-y\|$ for all $(x,y) \in \mathcal{H} \times \mathcal{H}$. A single-valued linear operator $M : \mathcal{H} \to \mathcal{H}$ is said to be *skew*, if $\langle x, Mx\rangle = 0$ for all $x \in \mathcal{H}$.

The *resolvent* and the *reflected resolvent* of an operator $M : \mathcal{H} \to 2^{\mathcal{H}}$ are

$$J_M = (\mathrm{Id} + M)^{-1} \text{ and } R_M = 2J_M - \mathrm{Id},$$

respectively, the operator Id denoting the identity on the underlying Hilbert space. When M is maximally monotone, its resolvent (and, consequently, its reflected resolvent) is a single-valued operator and, by [2, Proposition 23.18], we have for $\gamma \in \mathbb{R}_{++}$

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

Moreover, for $f \in \Gamma(\mathcal{H})$ and $\gamma \in \mathbb{R}_{++}$ the subdifferential $\partial(\gamma f)$ is maximally monotone (see [34]) 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 γf at $x \in \mathcal{H}$ and it represents the unique optimal solution of the optimization problem

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

In this particular situation, (1) becomes *Moreau's decomposition formula*

$$\mathrm{Id} = \mathrm{Prox}_{\gamma f} + \gamma \mathrm{Prox}_{\gamma^{-1} f^*} \circ \gamma^{-1}\mathrm{Id}. \tag{2}$$

When $\Omega \subseteq \mathcal{H}$ is a nonempty, convex and closed set, the function $\delta_\Omega : \mathcal{H} \to \overline{\mathbb{R}}$, defined by $\delta_\Omega(x) = 0$ for $x \in \Omega$ and $\delta_\Omega(x) = +\infty$, otherwise, denotes the *indicator function* of the set Ω. For each $\gamma > 0$ the proximal point of $\gamma\delta_\Omega$ at $x \in \mathcal{H}$ is nothing else than

$$\mathrm{Prox}_{\gamma\delta_\Omega}(x) = \mathrm{Prox}_{\delta_\Omega}(x) = \mathcal{P}_\Omega(x) = \arg\min_{y \in \Omega} \frac{1}{2}\|y-x\|^2,$$

where $\mathcal{P}_\Omega : \mathcal{H} \to \Omega$ denotes the *projection operator* on Ω.

Finally, the *parallel sum* of two set-valued operators $M_1, M_2 : \mathcal{H} \to 2^{\mathcal{H}}$ is defined as $M_1 \Box M_2 : \mathcal{H} \to 2^{\mathcal{H}}$, $M_1 \Box M_2 = \left(M_1^{-1} + M_2^{-1}\right)^{-1}$.

3 Primal-dual algorithms for monotone inclusion problems

The following monotone inclusion problem will be in the focus of our investigations.

Problem 1. Let $\mathscr{H}$ be a real Hilbert space, $z \in \mathscr{H}$, $A : \mathscr{H} \to 2^{\mathscr{H}}$ a maximally monotone operator and $C : \mathscr{H} \to \mathscr{H}$ a monotone operator. Let m be a strictly positive integer and for any $i = 1,\dots,m$, let $\mathscr{G}_i$ be a real Hilbert space, $r_i \in \mathscr{G}_i$, $B_i, D_i : \mathscr{G}_i \to 2^{\mathscr{G}_i}$ be maximally monotone operators and $L_i : \mathscr{H} \to \mathscr{G}_i$ a nonzero linear continuous operator. The problem is to solve the primal inclusion

$$\text{find } \overline{x} \in \mathscr{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^* \big((B_i \Box D_i)(L_i\overline{x} - r_i)\big) + C\overline{x}, \tag{3}$$

together with the dual inclusion of Attouch-Théra type (see [1, 23, 40])

$$\text{find } \overline{v}_1 \in \mathscr{G}_1, \dots, \overline{v}_m \in \mathscr{G}_m \text{ such that } \exists x \in \mathscr{H} : \begin{cases} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in (B_i \Box D_i)(L_i x - r_i),\ i = 1,\dots,m. \end{cases} \tag{4}$$

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

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in A\overline{x} + C\overline{x} \text{ and } \overline{v}_i \in (B_i \Box D_i)(L_i\overline{x} - r_i),\ i = 1,\dots,m. \tag{5}$$

If $\overline{x} \in \mathscr{H}$ is a solution to (3), then there exists $(\overline{v}_1, \dots, \overline{v}_m) \in \mathscr{G}_1 \times \dots \times \mathscr{G}_m$ such that $(\overline{x}, \overline{v}_1, \dots, \overline{v}_m)$ is a primal-dual solution to Problem 1 and, if $(\overline{v}_1, \dots, \overline{v}_m) \in \mathscr{G}_1 \times \dots \times \mathscr{G}_m$ is a solution to (4), then there exists $\overline{x} \in \mathscr{H}$ such that $(\overline{x}, \overline{v}_1, \dots, \overline{v}_m)$ is a primal-dual solution to Problem 1. Moreover, if $(\overline{x}, \overline{v}_1, \dots, \overline{v}_m) \in \mathscr{H} \times \mathscr{G}_1 \times \dots \times \mathscr{G}_m$ is a primal-dual solution to Problem 1, then $\overline{x}$ is a solution to (3) and $(\overline{v}_1, \dots, \overline{v}_m) \in \mathscr{G}_1 \times \dots \times \mathscr{G}_m$ is a solution to (4).

3.1 Forward-backward type algorithms

By employing the classical forward-backward algorithm (see [21, 39]) in an appropriate product space, Vũ proposed in [40] an iterative scheme for solving a slightly modified version of Problem 1 formulated in the presence of some given weights $w_i \in (0,1], i = 1,\dots,m$, 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 [40, Theorem 3.1] in the error-free case and when $\lambda_n = 1$ for any $n \geq 0$.

Theorem 1. *In Problem 1 suppose that C is η-cocoercive and D_i is ν_i-strongly monotone with $\eta, \nu_i > 0$ for $i = 1,\dots,m$. Moreover, assume 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, $i = 1,\dots,m$, be strictly positive numbers such that

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

Let $(x_0, v_{1,0}, \dots, v_{m,0}) \in \mathcal{H} \times \mathcal{G}_1 \times \dots \times \mathcal{G}_m$ and set:

$$(\forall n \geq 0) \left\lfloor \begin{array}{l} x_{n+1} = J_{\tau A}\left[x_n - \tau\left(\sum_{i=1}^{m} L_i^* v_{i,n} + Cx_n - z\right)\right] \\ y_n = 2x_{n+1} - x_n \\ v_{i,n+1} = J_{\sigma_i B_i^{-1}}[v_{i,n} + \sigma_i(L_i y_n - D_i^{-1} v_{i,n} - r_i)],\ i = 1,\dots,m. \end{array} \right.$$

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

In the remaining of this subsection we propose in two different settings modified versions of the algorithm in Theorem 1 and discuss the orders of convergence of the sequences of iterates generated by the new iterative schemes.

3.1.1 The case $A + C$ is strongly monotone

Additionally to the hypotheses in Problem 1 we assume throughout this subsection that

$$(H_1) \begin{cases} (i)\ A + C \text{ is } \gamma-\text{strongly monotone with } \gamma > 0; \\ (ii)\ D_i^{-1}(x) = 0 \text{ for all } x \in \mathcal{G}_i, i = 1,\dots,m; \\ (iii)\ C \text{ is } \eta-\text{Lipschitzian with } \ \eta > 0. \end{cases}$$

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 of primal iterates $(x_n)_{n \geq 0}$. To this end, we update in each iteration the parameters τ and σ_i, $i = 1,\dots,m$, and use a modified formula for the sequence $(y_n)_{n \geq 0}$. Due to technical reasons, we apply this method in the particular case stated by (ii) above. In the light of the approach described in Remark 3 below, one can extend the statement of Theorem 3, which is the convergence statement for the modified iterative scheme, to the primal-dual pair of monotone inclusions stated in Problem 1.

Remark 1. Different to the hypotheses of Problem 1, we relax the assumptions made on the operator C. It is obvious that, if C is a η-cocoercive operator with $\eta > 0$, 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, [2, 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, $(x,v) \mapsto (L^*v, -Lx)$ is an operator having all these properties. This operator appears in a natural way when considering primal-dual monotone inclusion problems as done in [17].

We propose the following modification of the iterative scheme in Theorem 1.

Algorithm 2. Let $(x_0, v_{1,0}, \dots, v_{m,0}) \in \mathscr{H} \times \mathscr{G}_1 \times \dots \times \mathscr{G}_m$, let $\tau_0 > 0$, $\sigma_{i,0} > 0$, $i = 1, \dots, m$, such that $\tau_0 < 2\gamma/\eta$, $\lambda \geq \eta + 1$, $\tau_0 \sum_{i=1}^m \sigma_{i,0} \|L_i\|^2 \leq \sqrt{1 + \tau_0(2\gamma - \eta\tau_0)/\lambda}$ and $\theta_0 = 1/\sqrt{1 + \tau_0(2\gamma - \eta\tau_0)/\lambda}$. Set

$$(\forall n \geq 0) \left\lfloor \begin{array}{l} x_{n+1} = J_{(\tau_n/\lambda)A}\left[x_n - (\tau_n/\lambda)\left(\sum_{i=1}^m L_i^* v_{i,n} + Cx_n - z\right)\right] \\ y_n = x_{n+1} + \theta_n(x_{n+1} - x_n) \\ v_{i,n+1} = J_{\sigma_{i,n} B_i^{-1}}[v_{i,n} + \sigma_{i,n}(L_i y_n - r_i)],\ i = 1, \dots, m \\ \tau_{n+1} = \theta_n \tau_n,\ \theta_{n+1} = 1/\sqrt{1 + \tau_{n+1}(2\gamma - \eta\tau_{n+1})/\lambda} \\ \sigma_{i,n+1} = \sigma_{i,n}/\theta_{n+1},\ i = 1, \dots, m. \end{array} \right.$$

Theorem 3. *In Problem 1 suppose that (H_1) holds and let $(\overline{x}, \overline{v}_1, \dots, \overline{v}_m)$ be a primal-dual solution to Problem 1. Then the sequences generated by Algorithm 2 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, $\lim\limits_{n\to+\infty} n\tau_n = \frac{\lambda}{\gamma}$, hence one obtains for $(x_n)_{n\geq 0}$ an order of convergence of $\mathscr{O}(\frac{1}{n})$.

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

$$\begin{aligned} &\frac{\lambda}{\tau_{n+2}^2}\|x_{n+2} - \overline{x}\|^2 + \sum_{i=1}^m \frac{\|v_{i,n+1} - \overline{v}_i\|^2}{\tau_1 \sigma_{i,0}} + \frac{\|x_{n+2} - x_{n+1}\|^2}{\tau_{n+1}^2} - \\ &\frac{2}{\tau_{n+1}} \sum_{i=1}^m \langle L_i(x_{n+2} - x_{n+1}), -v_{i,n+1} + \overline{v}_i \rangle \leq \\ &\frac{\lambda}{\tau_{n+1}^2}\|x_{n+1} - \overline{x}\|^2 + \sum_{i=1}^m \frac{\|v_{i,n} - \overline{v}_i\|^2}{\tau_1 \sigma_{i,0}} + \frac{\|x_{n+1} - x_n\|^2}{\tau_n^2} - \\ &\frac{2}{\tau_n} \sum_{i=1}^m \langle L_i(x_{n+1} - x_n), -v_{i,n} + \overline{v}_i \rangle. \end{aligned} \tag{6}$$

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

$$\frac{\lambda}{\tau_{n+1}}(x_{n+1}-x_{n+2}) - \left(\sum_{i=1}^m L_i^* v_{i,n+1} + Cx_{n+1} - z\right) + Cx_{n+2} \in (A+C)x_{n+2}. \quad (7)$$

Since $A+C$ is γ-strongly monotone, (5) and (7) yield for any $n \geq 0$

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

Further, we have

$$\langle x_{n+2}-\overline{x}, x_{n+1}-x_{n+2}\rangle = \frac{\|x_{n+1}-\overline{x}\|^2}{2} - \frac{\|x_{n+2}-\overline{x}\|^2}{2} - \frac{\|x_{n+1}-x_{n+2}\|^2}{2} \quad (8)$$

and, since C is η-Lipschitzian,

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

hence for any $n \geq 0$ it yields (taking into account that $\lambda \geq \eta + 1$)

$$\left(\frac{\lambda}{\tau_{n+1}} + 2\gamma - \eta\tau_{n+1}\right)\|x_{n+2}-\overline{x}\|^2 \leq \\ \frac{\lambda}{\tau_{n+1}}\|x_{n+1}-\overline{x}\|^2 - \frac{1}{\tau_{n+1}}\|x_{n+2}-x_{n+1}\|^2 + 2\sum_{i=1}^m \langle L_i(x_{n+2}-\overline{x}), \overline{v}_i - v_{i,n+1}\rangle. \quad (9)$$

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

$$\frac{1}{\sigma_{i,n}}(v_{i,n}-v_{i,n+1}) + L_i y_n - r_i \in B_i^{-1} v_{i,n+1}, \quad (10)$$

the monotonicity of B_i^{-1} and (5) we obtain

$$0 \leq \frac{1}{2\sigma_{i,n}}\|v_{i,n}-\overline{v}_i\|^2 - \frac{1}{2\sigma_{i,n}}\|v_{i,n}-v_{i,n+1}\|^2 - \frac{1}{2\sigma_{i,n}}\|v_{i,n+1}-\overline{v}_i\|^2 \\ + \langle L_i(y_n-\overline{x}), v_{i,n+1}-\overline{v}_i\rangle,$$

which yields (use also (9)) for any $n \geq 0$

$$\left(\frac{\lambda}{\tau_{n+1}} + 2\gamma - \eta\tau_{n+1}\right)\|x_{n+2}-\overline{x}\|^2 + \sum_{i=1}^m \frac{\|v_{i,n+1}-\overline{v}_i\|^2}{\sigma_{i,n}} \leq \\ \frac{\lambda}{\tau_{n+1}}\|x_{n+1}-\overline{x}\|^2 + \sum_{i=1}^m \frac{\|v_{i,n}-\overline{v}_i\|^2}{\sigma_{i,n}} - \frac{\|x_{n+2}-x_{n+1}\|^2}{\tau_{n+1}} - \sum_{i=1}^m \frac{\|v_{i,n}-v_{i,n+1}\|^2}{\sigma_{i,n}} \quad (11)$$

$$+2\sum_{i=1}^{m}\langle L_i(x_{n+2}-y_n),-v_{i,n+1}+\overline{v}_i\rangle.$$

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

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

By combining the last inequality with (11) we obtain for any $n\geq 0$

$$\begin{aligned}&\left(\frac{\lambda}{\tau_{n+1}}+2\gamma-\eta\tau_{n+1}\right)\|x_{n+2}-\overline{x}\|^2+\sum_{i=1}^{m}\frac{\|v_{i,n+1}-\overline{v}_i\|^2}{\sigma_{i,n}}+\frac{\|x_{n+2}-x_{n+1}\|^2}{\tau_{n+1}}-\\&2\sum_{i=1}^{m}\langle L_i(x_{n+2}-x_{n+1}),-v_{i,n+1}+\overline{v}_i\rangle\leq\\&\frac{\lambda}{\tau_{n+1}}\|x_{n+1}-\overline{x}\|^2+\sum_{i=1}^{m}\frac{\|v_{i,n}-\overline{v}_i\|^2}{\sigma_{i,n}}+\left(\sum_{i=1}^{m}\|L_i\|^2\sigma_{i,n}\right)\theta_n^2\|x_{n+1}-x_n\|^2-\\&2\sum_{i=1}^{m}\theta_n\langle L_i(x_{n+1}-x_n),-v_{i,n}+\overline{v}_i\rangle.\end{aligned}\tag{12}$$

After dividing (12) 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 (6) is true.

Let $N\in\mathbb{N},N\geq 2$. Summing up the inequality in (6) from $n=0$ to $N-1$, it yields

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

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

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

and obtain finally the inequality in the statement of the theorem.

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

$$\tau_{n+1} = \frac{\tau_n}{\sqrt{1+\frac{\tau_n}{\lambda}(2\gamma-\eta\tau_n)}}. \tag{13}$$

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 (13) one easily obtains that $\lim_{n\to+\infty} \tau_n = 0$ and, further, that $\lim_{n\to+\infty} \frac{\tau_n}{\tau_{n+1}} = 1$. As $(\frac{1}{\tau_n})_{n\geq 0}$ is a strictly increasing and unbounded sequence, by applying the Stolz-Cesàro Theorem, it yields (see [9])

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

$\square$

Remark 2. If $A+C$ is strongly monotone, 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 (3) has at most one solution. Hence, if $(\overline{x},\overline{v}_1,...,\overline{v}_m)$ is a primal-dual solution to Problem 1, then $\overline{x}$ is the unique solution to (3). Notice that the problem (4) may not have a unique solution.

Remark 3. In Algorithm 2 and Theorem 3 we assumed that $D_i^{-1}=0$ for $i=1,\dots,m$, however, similar statements can be also provided for Problem 1 under the additional assumption that the operators $D_i:\mathcal{G}_i\to 2^{\mathcal{G}_i}$ are ν_i^{-1}-cocoercive with $\nu_i\in\mathbb{R}_{++}$ for $i=1,\dots,m$. This assumption is in general stronger than assuming that D_i is monotone and D_i^{-1} is ν_i-Lipschitzian for $i=1,...,m$ and it guarantees that D_i is ν_i^{-1}-strongly monotone and maximally monotone for $i=1,...,m$ (see [2, Example 20.28, Proposition 20.22 and Example 22.6]). We introduce the Hilbert space $\widetilde{\mathcal{H}} = \mathcal{H}\times\widetilde{\mathcal{G}}$, where $\widetilde{\mathcal{G}} = \mathcal{G}_1\times...\times\mathcal{G}_m$, the element $\widetilde{z}=(z,0,\dots,0)\in\widetilde{\mathcal{H}}$ and the maximally monotone operator $\widetilde{A}:\widetilde{\mathcal{H}}\to 2^{\widetilde{\mathcal{H}}}$, $\widetilde{A}(x,y_1,\dots,y_m)=(Ax,D_1y_1,\dots,D_my_m)$ and the monotone and Lipschitzian operator $\widetilde{C}:\widetilde{\mathcal{H}}\to\widetilde{\mathcal{H}}$, $\widetilde{C}(x,y_1,\dots,y_m)=(Cx,0,\dots,0)$. Notice also that $\widetilde{A}+\widetilde{C}$ is strongly monotone. Furthermore, we introduce the element $\widetilde{r}=(r_1,\dots,r_m)\in\widetilde{\mathcal{G}}$, the maximally monotone operator $\widetilde{B}:\widetilde{\mathcal{G}}\to 2^{\widetilde{\mathcal{G}}}$, $\widetilde{B}(y_1,\dots,y_m)=(B_1y_1,\dots,B_my_m)$, and the linear continuous operator $\widetilde{L}:\widetilde{\mathcal{H}}\to\widetilde{\mathcal{G}}$, $\widetilde{L}(x,y_1\dots,y_m)=(L_1x-y_1,\dots,L_mx-y_m)$, having as adjoint $\widetilde{L}^*:\widetilde{\mathcal{G}}\to\widetilde{\mathcal{H}}$, $\widetilde{L}^*(q_1,\dots,q_m)=(\sum_{i=1}^m L_i^*q_i,-q_1,\dots,-q_m)$. We consider the primal problem

$$\text{find } \widetilde{x}=(\overline{x},\overline{p}_1\dots\overline{p}_m)\in\widetilde{\mathcal{H}} \text{ such that } \widetilde{z}\in\widetilde{A}\widetilde{x}+\widetilde{L}^*\widetilde{B}\left(\widetilde{L}\widetilde{x}-\widetilde{r}\right)+\widetilde{C}\widetilde{x}, \tag{14}$$

together with the dual inclusion problem

$$\text{find } \widetilde{v} \in \widetilde{\mathscr{G}} \text{ such that } \exists \widetilde{x} \in \widetilde{\mathscr{H}} : \begin{cases} \widetilde{z} - \widetilde{L}^* \widetilde{v} \in \widetilde{A}\widetilde{x} + \widetilde{C}\widetilde{x} \\ \widetilde{v} \in \widetilde{B}(\widetilde{L}\widetilde{x} - \widetilde{r}) \end{cases}. \tag{15}$$

We notice that Algorithm 2 can be employed for solving this primal-dual pair of monotone inclusion problems and that in its formulation the resolvents of A, B_i and $D_i, i = 1,...,m$ are separately involved, as for $\gamma \in \mathbb{R}_{++}$

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

We have that $(\widetilde{x}, \widetilde{v})$ is a a primal-dual solution to (14)-(15) if and only if

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

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

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

Thus, if $(\widetilde{x}, \widetilde{v})$ is a primal-dual solution to (14)-(15), then $(\overline{x}, \widetilde{v})$ is a primal-dual solution to Problem 1. Viceversa, if $(\overline{x}, \widetilde{v})$ is a primal-dual solution to Problem 1, then, choosing $\overline{p}_i \in D_i^{-1}\overline{v}_i, i = 1,...,m$, and $\widetilde{x} = (\overline{x}, \overline{p}_1 \ldots \overline{p}_m)$, it yields that $(\widetilde{x}, \widetilde{v})$ is a primal-dual solution to (14)-(15). In conclusion, the first component of every primal iterate in $\widetilde{\mathscr{H}}$ generated by Algorithm 2 for finding a primal-dual solution $(\widetilde{x}, \widetilde{v})$ to (14)-(15) will furnish a sequence of iterates in $\mathscr{H}$ fulfilling the inequality in the formulation of Theorem 3 for the primal-dual solution $(\overline{x}, \widetilde{v})$ to Problem 1.

3.1.2 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 1 which guarantees, when $A + C$ and $B_i^{-1} + D_i^{-1}, i = 1,...,m$, are strongly monotone, orders of convergence of $\mathscr{O}(\omega^n)$, for $\omega \in (0,1)$, for the sequences of iterates $(x_n)_{n\geq 0}$ 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 the following hypotheses

$$(H_2) \begin{cases} (i)\; A + C \text{ is } \gamma-\text{strongly monotone with } \gamma > 0; \\ (ii)\; B_i^{-1} + D_i^{-1} \text{ is } \delta_i-\text{strongly monotone with } \delta_i > 0, i = 1,....m; \\ (iii)\; D_i^{-1} \text{ is } \nu_i-\text{Lipschitzian with } \nu_i > 0, i = 1,...,m; \\ (iv)\; C \text{ is } \eta-\text{Lipschitzian with } \; \eta > 0. \end{cases}$$

We propose the following modification of the iterative scheme in Theorem 1.

Algorithm 4. Let $(x_0, v_{1,0}, ..., v_{m,0}) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$, let $\mu > 0$ such that $\mu \leq \min\left\{\gamma^2/\eta^2, \delta_1^2/\nu_1^2, \dots, \delta_m^2/\nu_m^2, \sqrt{\gamma/\left(\sum_{i=1}^m \|L_i\|^2/\delta_i\right)}\right\}$, $\tau = \mu/(2\gamma)$, $\sigma_i = \mu/(2\delta_i)$, $i = 1, ..., m$, and $\theta \in [2/(2+\mu), 1]$. Set

$$(\forall n \geq 0) \left\lfloor \begin{array}{l} x_{n+1} = J_{\tau A}\left[x_n - \tau\left(\sum_{i=1}^m L_i^* v_{i,n} + Cx_n - z\right)\right] \\ y_n = x_{n+1} + \theta(x_{n+1} - x_n) \\ v_{i,n+1} = J_{\sigma_i B_i^{-1}}[v_{i,n} + \sigma_i(L_i y_n - D_i^{-1} v_{i,n} - r_i)], \ i = 1, ..., m. \end{array} \right.$$

For the proof of the following result we refer to [9].

Theorem 5. *In Problem 1 suppose that* (H_2) *holds and let* $(\bar{x}, \bar{v}_1, ..., \bar{v}_m)$ *be a primal-dual solution to Problem 1. Then the sequences generated by Algorithm 4 fulfill for any* $n \geq 0$

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

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

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

Remark 4. If $A + C$ and $B_i^{-1} + D_i^{-1}$ are strongly monotone $i = 1, ..., m$, then there exists at most one primal-dual solution to Problem 1. Hence, if $(\bar{x}, \bar{v}_1, ..., \bar{v}_m)$ is a primal-dual solution to Problem 1, then $\bar{x}$ is the unique solution to the primal inclusion (3) and $(\bar{v}_1, ..., \bar{v}_m)$ is the unique solution to the dual inclusion (4).

3.2 Forward-backward-forward type algorithms

In this subsection we recall the error-free variant of the primal-dual algorithm in [23] and the corresponding convergence statements, as given in [23, Theorem 3.1], and propose two accelerated versions of it. The proof of the following statement relies on the application of the error Tseng's forward-backward-forward algorithm in a product space.

Theorem 6. *In Problem 1 suppose that* C *is* μ*-Lipschitian with* $\mu > 0$ *and* D_i^{-1} *is* ν_i*-Lipschitzian with* $\nu_i > 0$ *for* $i = 1, ..., m$*. Moreover, assume that*

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

Let $x_0 \in \mathcal{H}$ *and* $(v_{1,0}, \dots, v_{m,0}) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$*, set* $\beta = \max\{\mu, \nu_1 \dots, \nu_m\} + \sqrt{\sum_{i=1}^m \|L_i\|^2}$*, choose* $\varepsilon \in (0, \frac{1}{\beta+1})$ *and* $(\gamma_n)_{n\geq 0}$ *a sequence in* $\left[\varepsilon, \frac{1-\varepsilon}{\beta}\right]$ *and set*

$$
(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} p_{1,n} = J_{\gamma_n A}\left(x_n - \gamma_n \left(Cx_n + \sum_{i=1}^m L_i^* v_{i,n} - z\right)\right) \\ \textit{For } i = 1, \ldots, m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = J_{\gamma_n B_i^{-1}}\left(v_{i,n} + \gamma_n (L_i x_n - D_i^{-1} v_{i,n} - r_i)\right) \\ v_{i,n+1} = \gamma_n L_i (p_{1,n} - x_n) + \gamma_n (D_i^{-1} v_{i,n} - D_i^{-1} p_{2,i,n}) + p_{2,i,n} \end{array} \right. \\ x_{n+1} = \gamma_n \sum_{i=1}^m L_i^* (v_{i,n} - p_{2,i,n}) + \gamma_n (Cx_n - Cp_{1,n}) + p_{1,n}. \end{array} \right. \tag{16}
$$

Then there exists a primal-dual solution $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$ *to Problem 1 such that* $x_n \rightharpoonup \bar{x}$, $p_{1,n} \rightharpoonup \bar{x}$, $(v_{1,n}, \ldots, v_{m,n}) \rightharpoonup (\bar{v}_1, \ldots, \bar{v}_m)$ *and* $(p_{2,1,n}, \ldots, p_{2,m,n}) \rightharpoonup (\bar{v}_1, \ldots, \bar{v}_m)$ *as* $n \to +\infty$.

3.2.1 The case $A + C$ is strongly monotone

Additionally to the hypotheses mentioned in Problem 1 we assume throughout this subsection that

$$
(H_3) \begin{cases} (i)\ A + C \text{ is } \rho-\text{strongly monotone with } \rho > 0; \\ (ii)\ D_i^{-1}(x) = 0 \text{ for all } x \in \mathcal{G}_i, i = 1, \ldots, m; \\ (iii)\ C \text{ is } \mu-\text{Lipschitzian with } \mu > 0. \end{cases}
$$

We refer to Remark 3 for how to handle the general Problem 1 in the situation when the operators D_i are involved, as well. The subsequent algorithm represents an accelerated version of the one given in Theorem 6 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.

Algorithm 7. Let $x_0 \in \mathcal{H}$, $(v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$,

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

Consider the following updates:

$$
(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} p_{1,n} = J_{\gamma_n A}\left(x_n - \gamma_n \left(Cx_n + \sum_{i=1}^m L_i^* v_{i,n} - z\right)\right) \\ \text{For } i = 1, \ldots, m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = J_{\sigma_n B_i^{-1}}\left(v_{i,n} + \sigma_n (L_i x_n - r_i)\right) \\ v_{i,n+1} = \sigma_n L_i (p_{1,n} - x_n) + p_{2,i,n} \end{array} \right. \\ x_{n+1} = \gamma_n \sum_{i=1}^m L_i^* (v_{i,n} - p_{2,i,n}) + \gamma_n (Cx_n - Cp_{1,n}) + p_{1,n} \\ \theta_n = 1/\sqrt{1 + 2\rho\gamma_n(1-\gamma_n)},\ \gamma_{n+1} = \theta_n \gamma_n,\ \sigma_{n+1} = \sigma_n/\theta_n. \end{array} \right. \tag{17}
$$

For the proof of the following convergence theorem we refer the reader to [13].

Theorem 8. *In Problem 1 suppose that* (H_3) *holds and let* $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$ *be a primal-dual solution to Problem 1. Then the sequences generated by Algorithm 7 satisfy for every* $n \geq 0$ *the inequality*

$$\|x_n - \overline{x}\|^2 + \gamma_n \sum_{i=1}^m \frac{\|v_{i,n} - \overline{v}_i\|^2}{\sigma_n} \leq \gamma_n^2 \left(\frac{\|x_0 - \overline{x}\|^2}{\gamma_0^2} + \sum_{i=1}^m \frac{\|v_{i,0} - \overline{v}_i\|^2}{\gamma_0 \sigma_0} \right). \tag{18}$$

Moreover, $\lim_{n \to +\infty} n\gamma_n = \frac{1}{\rho}$*, hence one obtains for* $(x_n)_{n \geq 0}$ *an order of convergence of* $\mathcal{O}(\frac{1}{n})$.

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

Assuming the hypotheses

$$(H_4) \begin{cases} (i)\ A+C \text{ is } \rho-\text{strongly monotone with } \gamma > 0; \\ (ii)\ B_i^{-1} + D_i^{-1} \text{ is } \tau_i-\text{strongly monotone with } \tau_i > 0, i = 1,....m; \\ (iii)\ D_i^{-1} \text{ is } \nu_i-\text{Lipschitzian with } \nu_i > 0, i = 1,...,m; \\ (iv)\ C \text{ is } \mu-\text{Lipschitzian with } \ \mu > 0 \end{cases}$$

fulfilled, we provide as follows a second accelerated version of the algorithm in Theorem 6 which generates sequences of primal and dual iterates that converge to the primal-dual solution to Problem 1 with an improved rate of convergence.

Algorithm 9. Let $x_0 \in \mathcal{H}$, $(v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$, 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:

$$(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} p_{1,n} = J_{\gamma A}\left(x_n - \gamma(Cx_n + \sum_{i=1}^m L_i^* v_{i,n} - z)\right) \\ \text{For } i = 1, \ldots, m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = J_{\gamma B_i^{-1}}\left(v_{i,n} + \gamma(L_i x_n - D_i^{-1} v_{i,n} - r_i)\right) \\ v_{i,n+1} = \gamma L_i(p_{1,n} - x_n) + \gamma(D_i^{-1} v_{i,n} - D_i^{-1} p_{2,i,n}) + p_{2,i,n} \end{array} \right. \\ x_{n+1} = \gamma \sum_{i=1}^m L_i^*(v_{i,n} - p_{2,i,n}) + \gamma(Cx_n - Cp_{1,n}) + p_{1,n}. \end{array} \right. \tag{19}$$

Theorem 10. *In Problem 1 suppose* (H_4) *holds and let* $(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ *be a primal-dual solution to Problem 1. Then the sequences generated by Algorithm 9 satisfy for every* $n \geq 0$ *the inequality*

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

where $\rho_{\min} = \min\{\rho, \tau_1, \ldots, \tau_m\}$.

Proof. Taking into account the definitions of the resolvents occurring in the iterative scheme of Algorithm 9, we obtain for every $n \geq 0$

$$\frac{x_n - x_{n+1}}{\gamma} - \sum_{i=1}^{m} L_i^* p_{2,i,n} + z \in (A+C)p_{1,n}$$

and

$$\frac{v_{i,n} - v_{i,n+1}}{\gamma} + L_i p_{1,n} - r_i \in (B_i^{-1} + D_i^{-1})p_{2,i,n},\ i = 1,\ldots,m.$$

By the strong monotonicity of $A+C$ and $B_i^{-1}+D_i^{-1}$, $i=1,\ldots,m$, from (5) we obtain for every $n \geq 0$

$$\left\langle p_{1,n} - \overline{x}, \frac{x_n - x_{n+1}}{\gamma} - \sum_{i=1}^{m} L_i^* p_{2,i,n} + z - \left(z - \sum_{i=1}^{m} L_i^* \overline{v}_i\right)\right\rangle \geq \rho \|p_{1,n} - \overline{x}\|^2 \tag{20}$$

and, respectively,

$$\left\langle p_{2,i,n} - \overline{v}_i, \frac{v_{i,n} - v_{i,n+1}}{\gamma} + L_i p_{1,n} - r_i - (L_i\overline{x} - r_i)\right\rangle \geq \tau_i \|p_{2,i,n} - \overline{v}_i\|^2,\ i = 1,...,m. \tag{21}$$

Consider the Hilbert space $\widetilde{\mathscr{H}} = \mathscr{H} \times \mathscr{G}_1 \times ... \times \mathscr{G}_m$, equipped with the usual inner product and associated norm, and set

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

Summing up the inequalities (20) and (21) and using

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

we obtain for every $n \geq 0$

$$\frac{\|\widetilde{x}_n - \widetilde{x}\|^2}{2\gamma} \geq \rho_{\min}\|\widetilde{p}_n - \widetilde{x}\|^2 + \frac{\|\widetilde{x}_{n+1} - \widetilde{x}\|^2}{2\gamma} + \frac{\|\widetilde{x}_n - \widetilde{p}_n\|^2}{2\gamma} - \frac{\|\widetilde{x}_{n+1} - \widetilde{p}_n\|^2}{2\gamma}. \tag{22}$$

Further, we obtain

$$\rho_{\min}\|\widetilde{p}_n - \widetilde{x}\|^2 \geq \frac{2\rho_{\min}\gamma(1-\gamma)}{2\gamma}\|\widetilde{x}_{n+1} - \widetilde{x}\|^2 - \frac{2\rho_{\min}}{2\gamma}\|\widetilde{x}_{n+1} - \widetilde{p}_n\|^2\ \forall n \geq 0.$$

Hence, from (22) we get for all $n \geq 0$

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

Further, we have for every $n \geq 0$ (see [13])

$$\frac{\|\widetilde{x}_n - \widetilde{p}_n\|^2}{2\gamma} - \frac{(1+2\rho_{\min})\|\widetilde{x}_{n+1} - \widetilde{p}_n\|^2}{2\gamma} \geq 0,$$

therefore, we obtain

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

which leads to

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

3.3 Douglas-Rachford type algorithms

The third class of primal-dual methods for solving Problem 1 that we discuss in this paper is the one of Douglas-Rachford type algorithms which was introduced in [14]. It has the particularity that the operators occurring in the parallel sums can be arbitrary maximally monotone ones, however, provided that $Cx = 0$ for all $x \in \mathscr{H}$.

3.3.1 A first Douglas-Rachford type primal-dual algorithm

The first iterative scheme of Douglas-Rachford type we deal with has the particularity that the operators A, B_i^{-1} and D_i^{-1}, $i = 1,...,m$, are accessed via their resolvents and that it processes each operator L_i and its adjoint L_i^*, $i = 1,...,m$, two times.

Algorithm 11. Let $x_0 \in \mathscr{H}$, $(v_{1,0},\dots,v_{m,0}) \in \mathscr{G}_1 \times \dots \times \mathscr{G}_m$ and τ and $\sigma_i, i = 1,...,m$, be strictly positive real numbers such that $\tau\sum_{i=1}^m \sigma_i\|L_i\|^2 < 4$. Furthermore, let $(\lambda_n)_{n\geq 0}$ be a sequence in $(0,2)$ and set

$$(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} p_{1,n} = J_{\tau A}\left(x_n - \frac{\tau}{2}\sum_{i=1}^m L_i^* v_{i,n} + \tau z\right) \\ w_{1,n} = 2p_{1,n} - x_n \\ \text{For } i = 1,\dots,m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = J_{\sigma_i B_i^{-1}}\left(v_{i,n} + \frac{\sigma_i}{2} L_i w_{1,n} - \sigma_i r_i\right) \\ w_{2,i,n} = 2p_{2,i,n} - v_{i,n} \end{array} \right. \\ z_{1,n} = w_{1,n} - \frac{\tau}{2}\sum_{i=1}^m L_i^* w_{2,i,n} \\ x_{n+1} = x_n + \lambda_n(z_{1,n} - p_{1,n}) \\ \text{For } i = 1,\dots,m \\ \quad \left\lfloor \begin{array}{l} z_{2,i,n} = J_{\sigma_i D_i^{-1}}\left(w_{2,i,n} + \frac{\sigma_i}{2} L_i(2z_{1,n} - w_{1,n})\right) \\ v_{i,n+1} = v_{i,n} + \lambda_n(z_{2,i,n} - p_{2,i,n}). \end{array} \right. \end{array} \right. \tag{23}$$

Theorem 12. *For Problem 1 assume that $Cx = 0$ for all $x \in \mathscr{H}$,*

$$z \in \operatorname{ran}\left(A + \sum_{i=1}^{m} L_i^*(B_i \Box D_i)(L_i \cdot - r_i)\right) \tag{24}$$

and consider the sequences generated by Algorithm 11.

1. *If* $\sum_{n=0}^{+\infty} \lambda_n(2-\lambda_n) = +\infty$, *then*
   a. $(x_n, v_{1,n}, \ldots, v_{m,n})_{n\geq 0}$ *converges weakly to a point* $(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$ *such that, when setting*

$$\overline{p}_1 = J_{\tau A}\left(\overline{x} - \frac{\tau}{2}\sum_{i=1}^{m} L_i^* \overline{v}_i + \tau z\right),$$
$$\text{and } \overline{p}_{2,i} = J_{\sigma_i B_i^{-1}}\left(\overline{v}_i + \frac{\sigma_i}{2} L_i(2\overline{p}_1 - \overline{x}) - \sigma_i r_i\right),\ i = 1,...,m,$$

   the element $(\overline{p}_1, \overline{p}_{2,1}, \ldots, \overline{p}_{2,m})$ *is a primal-dual solution to Problem 1.*
   b. $\lambda_n(z_{1,n} - p_{1,n}) \to 0$ $(n \to +\infty)$ *and* $\lambda_n(z_{2,i,n} - p_{2,i,n}) \to 0$ $(n \to +\infty)$ *for* $i = 1,...,m$.
   c. *whenever* $\mathcal{H}$ *and* $\mathcal{G}_i$, $i = 1,...,m$, *are finite-dimensional Hilbert spaces,* $(p_{1,n}, p_{2,1,n}, \ldots, p_{2,m,n})_{n\geq 0}$ *converges strongly to a primal-dual solution to Problem 1.*
2. *If* $\inf_{n\geq 0} \lambda_n > 0$ *and*

$$A \text{ and } B_i^{-1},\ i = 1,...,m, \text{ are uniformly monotone},$$

   then $(p_{1,n}, p_{2,1,n}, \ldots, p_{2,m,n})_{n\geq 0}$ *converges strongly to the unique primal-dual solution to Problem 1.*

Proof. Consider the Hilbert space $\mathcal{G} = \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$ endowed with inner product and associated norm defined, for $v = (v_1, \ldots, v_m)$, $q = (q_1, \ldots, q_m) \in \mathcal{G}$, as

$$\langle v, q\rangle = \sum_{i=1}^{m} \langle v_i, q_i\rangle \text{ and } \|v\| = \sqrt{\sum_{i=1}^{m} \|v_i\|^2}, \tag{25}$$

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

$$\langle (x,v),(y,q)\rangle = \langle x,y\rangle + \langle v,q\rangle \text{ and } \|(x,v)\| = \sqrt{\|x\|^2 + \|v\|^2}, \tag{26}$$

respectively. Consider the maximally monotone operator

$$M : \mathcal{K} \to 2^{\mathcal{K}}, \quad (x, v_1, \ldots, v_m) \mapsto (-z + Ax, r_1 + B_1^{-1} v_1, \ldots, r_m + B_m^{-1} v_m),$$

and the linear continuous operator

$$S : \mathscr{K} \to \mathscr{K}, \quad (x, v_1, \ldots, v_m) \mapsto \left(\sum_{i=1}^m L_i^* v_i, -L_1 x, \ldots, -L_m x \right),$$

which proves to be skew (i. e. $S^* = -S$) and hence maximally monotone (cf. [2, Example 20.30]). Further, consider the maximally monotone operator

$$Q : \mathscr{K} \to 2^{\mathscr{K}}, \quad (x, v_1, \ldots, v_m) \mapsto \left(0, D_1^{-1} v_1, \ldots, D_m^{-1} v_m\right).$$

Since $\operatorname{dom} S = \mathscr{K}$, both $\frac{1}{2}S + Q$ and $\frac{1}{2}S + M$ are maximally monotone (cf. [2, Corollary 24.4(i)]). On the other hand, according to [23, Eq. (3.12)], it holds (24) $\Leftrightarrow$ $\operatorname{zer}(M + S + Q) \neq \varnothing$, while [23, Eq. (3.21) and (3.22)] yield

$$(x, v_1, \ldots, v_m) \in \operatorname{zer}(M + S + Q) \Rightarrow (x, v_1, \ldots, v_m) \text{ is a primal-dual solution to Problem 1.} \tag{27}$$

Finally, we introduce the linear continuous operator

$$V : \mathscr{K} \to \mathscr{K},\ (x, v_1, \ldots, v_m) \quad \mapsto \left(\frac{x}{\tau} - \frac{1}{2} \sum_{i=1}^m L_i^* v_i, \frac{v_1}{\sigma_1} - \frac{1}{2} L_1 x, \ldots, \frac{v_m}{\sigma_m} - \frac{1}{2} L_m x \right).$$

It is a simple calculation to prove that V is self-adjoint, i. e. $V^* = V$. Furthermore, the operator V is ρ-strongly positive (see [14]) for

$$\rho = \left(1 - \frac{1}{2}\sqrt{\tau \sum_{i=1}^m \sigma_i \|L_i\|^2}\right) \min\left\{\frac{1}{\tau}, \frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_m}\right\},$$

which is a positive real number due to the assumptions made. Indeed for each $i = 1, \ldots, m$

$$2\|L_i\|\|x\|\|v_i\| \leq \frac{\sigma_i \|L_i\|^2}{\sqrt{\tau \sum_{i=1}^m \sigma_i \|L_i\|^2}} \|x\|^2 + \frac{\sqrt{\tau \sum_{i=1}^m \sigma_i \|L_i\|^2}}{\sigma_i} \|v_i\|^2 \tag{28}$$

and, consequently, for each $\widetilde{x} = (x, v_1, \ldots, v_m) \in \mathscr{K}$, it follows that

$$\begin{aligned}
\langle \widetilde{x}, V\widetilde{x} \rangle &\geq \frac{\|x\|^2}{\tau} + \sum_{i=1}^m \frac{\|v_i\|^2}{\sigma_i} - \sum_{i=1}^m \|L_i\|\|x\|\|v_i\| \\
&\geq \left(1 - \frac{1}{2}\sqrt{\tau \sum_{i=1}^m \sigma_i \|L_i\|^2}\right) \left(\frac{\|x\|^2}{\tau} + \sum_{i=1}^m \frac{\|v_i^2\|}{\sigma_i}\right) \\
&\geq \left(1 - \frac{1}{2}\sqrt{\tau \sum_{i=1}^m \sigma_i \|L_i\|^2}\right) \min\left\{\frac{1}{\tau}, \frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_m}\right\} \|\widetilde{x}\|^2 \\
&= \rho \|\widetilde{x}\|^2.
\end{aligned} \tag{29}$$

Since V is ρ-strongly positive, we have $\mathrm{cl}(\mathrm{ran}\,V) = \mathrm{ran}\,V$ (cf. [2, Fact 2.19]), $\mathrm{zer}\,V = \{0\}$ and, as $(\mathrm{ran}\,V)^{\perp} = \mathrm{zer}\,V^* = \mathrm{zer}\,V = \{0\}$ (see, for instance, [2, Fact 2.18]), it holds $\mathrm{ran}\,V = \mathscr{K}$. Consequently, V^{-1} exists and $\|V^{-1}\| \leq \frac{1}{\rho}$.

The algorithmic scheme (23) is equivalent to

$$(\forall n \geq 0) \left\lfloor \begin{array}{l} \frac{x_n - p_{1,n}}{\tau} - \frac{1}{2}\sum_{i=1}^m L_i^* v_{i,n} \in A p_{1,n} - z \\ w_{1,n} = 2p_{1,n} - x_n \\ \text{For } i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} \frac{v_{i,n} - p_{2,i,n}}{\sigma_i} - \frac{1}{2} L_i(x_n - p_{1,n}) \in -\frac{1}{2} L_i p_{1,n} + B_i^{-1} p_{2,i,n} + r_i \\ w_{2,i,n} = 2p_{2,i,n} - v_{i,n} \end{array} \right. \\ \frac{w_{1,n} - z_{1,n}}{\tau} - \frac{1}{2}\sum_{i=1}^m L_i^* w_{2,i,n} = 0 \\ x_{n+1} = x_n + \lambda_n (z_{1,n} - p_{1,n}) \\ \text{For } i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} \frac{w_{2,i,n} - z_{2,i,n}}{\sigma_i} - \frac{1}{2} L_i(w_{1,n} - z_{1,n}) \in -\frac{1}{2} L_i z_{1,n} + D_i^{-1} z_{2,i,n} \\ v_{i,n+1} = v_{i,n} + \lambda_n (z_{2,i,n} - p_{2,i,n}). \end{array} \right. \end{array} \right. \tag{30}$$

We introduce for every $n \geq 0$ the following notations:

$$\begin{cases} \widetilde{x}_n = (x_n, v_{1,n}, \ldots, v_{m,n}) \\ \widetilde{y}_n = (p_{1,n}, p_{2,1,n}, \ldots, p_{2,m,n}) \\ \widetilde{w}_n = (w_{1,n}, w_{2,1,n}, \ldots, w_{2,m,n}) \\ \widetilde{z}_n = (z_{1,n}, z_{2,1,n}, \ldots, z_{2,m,n}) \end{cases}. \tag{31}$$

The scheme (30) can equivalently be written in the form

$$(\forall n \geq 0) \left\lfloor \begin{array}{l} V(\widetilde{x}_n - \widetilde{y}_n) \in \left(\frac{1}{2}S + M\right)\widetilde{y}_n \\ \widetilde{w}_n = 2\widetilde{y}_n - \widetilde{x}_n \\ V(\widetilde{w}_n - \widetilde{z}_n) \in \left(\frac{1}{2}S + Q\right)\widetilde{z}_n \\ \widetilde{x}_{n+1} = \widetilde{x}_n + \lambda_n (\widetilde{z}_n - \widetilde{y}_n). \end{array} \right. \tag{32}$$

Next we introduce the Hilbert space $\mathscr{K}_V$ with inner product and norm respectively defined, for $\widetilde{x}, \widetilde{y} \in \mathscr{K}$, as

$$\langle \widetilde{x}, \widetilde{y} \rangle_{\mathscr{K}_V} = \langle \widetilde{x}, V\widetilde{y} \rangle \text{ and } \|\widetilde{x}\|_{\mathscr{K}_V} = \sqrt{\langle \widetilde{x}, V\widetilde{x} \rangle}, \tag{33}$$

respectively. Since $\frac{1}{2}S + M$ and $\frac{1}{2}S + Q$ are maximally monotone on $\mathscr{K}$, the operators

$$B := V^{-1}\left(\frac{1}{2}S + M\right) \text{ and } A := V^{-1}\left(\frac{1}{2}S + Q\right) \tag{34}$$

are maximally monotone on $\mathscr{K}_V$. Moreover, since V is self-adjoint and ρ-strongly positive, one can easily see that weak and strong convergence in $\mathscr{K}_V$ are equivalent with weak and strong convergence in $\mathscr{K}$, respectively.

Consequently, for every $n \geq 0$ we have (see [14])

$$
\begin{aligned}
&V(\widetilde{x}_n-\widetilde{y}_n)\in\left(\frac{1}{2}S+M\right)\widetilde{y}_n \Leftrightarrow V\widetilde{x}_n\in\left(V+\frac{1}{2}S+M\right)\widetilde{y}_n\\
&\Leftrightarrow \widetilde{x}_n\in\left(\mathrm{Id}+V^{-1}\left(\frac{1}{2}S+M\right)\right)\widetilde{y}_n \Leftrightarrow \widetilde{y}_n=\left(\mathrm{Id}+V^{-1}\left(\frac{1}{2}S+M\right)\right)^{-1}\widetilde{x}_n\\
&\Leftrightarrow \widetilde{y}_n=(\mathrm{Id}+B)^{-1}\widetilde{x}_n \qquad (35)
\end{aligned}
$$

and

$$
\begin{aligned}
&V(\widetilde{w}_n-\widetilde{z}_n)\in\left(\frac{1}{2}S+Q\right)\widetilde{z}_n \Leftrightarrow \widetilde{z}_n=\left(\mathrm{Id}+V^{-1}\left(\frac{1}{2}S+Q\right)\right)^{-1}\widetilde{w}_n\\
&\Leftrightarrow \widetilde{z}_n=(\mathrm{Id}+A)^{-1}\widetilde{w}_n. \qquad (36)
\end{aligned}
$$

Thus, the iterative rules in (32) become

$$
(\forall n\geq 0)\ \left\lfloor\begin{array}{l}\widetilde{y}_n=J_B\widetilde{x}_n\\ \widetilde{z}_n=J_A(2\widetilde{y}_n-\widetilde{x}_n)\\ \widetilde{x}_{n+1}=\widetilde{x}_n+\lambda_n(\widetilde{z}_n-\widetilde{y}_n)\end{array}\right., \qquad (37)
$$

which is nothing else than the error-free Douglas-Rachford algorithm (see [22]).

1. We assume that $\sum_{n=0}^{+\infty}\lambda_n(2-\lambda_n)=+\infty$ and are going to prove the statements in the first item.

1.a. According to [22, Theorem 2.1(i)(a)] the sequence $(\widetilde{x}_n)_{n\geq 0}$ converges weakly in $\mathcal{K}_V$ and, consequently, in $\mathcal{K}$ to a point $\overline{x}\in\operatorname{fix}(R_AR_B)$ with $J_B\overline{x}\in\operatorname{zer}(A+B)$. The claim follows by identifying $J_B\overline{x}$ and by noting (27).

1.b. According to [22, Theorem 2.1(i)(b)] it follows that $(R_AR_B\widetilde{x}_n-\widetilde{x}_n)\to 0$ $(n\to+\infty)$. The claim follows by taking into account that for every $n\geq 0$

$$
\lambda_n(z_n-y_n)=\frac{\lambda_n}{2}\left(R_AR_B\widetilde{x}_n-\widetilde{x}_n\right).
$$

1.c. As shown in a., we have that $\widetilde{x}_n\to\overline{x}\in\operatorname{fix}(R_AR_B)$ $(n\to+\infty)$ with $J_B\overline{x}\in\operatorname{zer}(A+B)=\operatorname{zer}(M+S+Q)$. Hence, by the continuity of J_B, we have

$$
\widetilde{y}_n=J_B\widetilde{x}_n\to J_B\overline{x}\in\operatorname{zer}(M+S+Q)\ (n\to+\infty).
$$

2. Assume that $\inf_{n\geq 0}\lambda_n>0$ and A and $B_i^{-1}, i=1,\dots,m$, are uniformly monotone. Then there exist increasing functions $\phi_A:\mathbb{R}_+\to[0,+\infty]$ and $\phi_{B_i^{-1}}:\mathbb{R}_+\to[0,+\infty], i=1,\dots,m$, vanishing only at 0, such that

$$
\begin{aligned}
&\langle x-y,u-z\rangle\geq\phi_A\left(\|x-y\|_{\mathcal{H}}\right)\ \forall(x,u),(y,z)\in\operatorname{gra}A\\
&\langle v-w,p-q\rangle\geq\phi_{B_i^{-1}}\left(\|v-w\|_{\mathcal{G}_i}\right)\ \forall(v,p),(w,q)\in\operatorname{gra}B_i^{-1}\ \forall i=1,\dots,m.
\end{aligned} \qquad (38)
$$

The function $\phi_M:\mathbb{R}_+\to[0,+\infty]$,

$$\phi_M(c) = \inf\left\{\phi_A(a) + \sum_{i=1}^{m} \phi_{B_i^{-1}}(b_i) : \sqrt{a^2 + \sum_{i=1}^{m} b_i^2} = c\right\}, \tag{39}$$

is increasing and vanishes only at 0 and it fulfills for each $(x,u),(y,z) \in \operatorname{gra} M$

$$\langle x-y, u-z\rangle_{\mathcal{K}} \geq \phi_M\left(\|x-y\|_{\mathcal{K}}\right). \tag{40}$$

Thus, M is uniformly monotone on $\mathcal{K}$.

The function $\phi_B : \mathbb{R}_+ \to [0,+\infty]$, $\phi_B(t) = \phi_M\left(\frac{1}{\sqrt{\|V\|}}t\right)$, is increasing and vanishes only at 0. Let be $(x,u),(y,z) \in \operatorname{gra} B$. Then there exist $v \in Mx$ and $w \in My$ fulfilling $Vu = \frac{1}{2}Sx + v$ and $Vz = \frac{1}{2}Sy + w$ and it holds (see [14])

$$\begin{aligned}\langle x-y, u-z\rangle_{\mathcal{K}_V} &\geq \phi_M\left(\frac{1}{\sqrt{\|V\|}}\|x-y\|_{\mathcal{K}_V}\right)\\ &\geq \phi_B\left(\|x-y\|_{\mathcal{K}_V}\right).\end{aligned} \tag{41}$$

Consequently, B is uniformly monotone on $\mathcal{K}_V$ and, according to [22, Theorem 2.1(ii)(b)], $(J_B x_n)_{n\geq 0}$ converges strongly to the unique element $\overline{y} \in \operatorname{zer}(A+B) = \operatorname{zer}(M+S+Q)$. Thus, $\widetilde{y}_n \to \overline{y}$ as $n \to +\infty$. $\square$

Remark 5. In the following we emphasize the relations between the proposed algorithm and other existent primal-dual iterative schemes.

(i) Other iterative methods for solving the primal-dual monotone inclusion pair introduced in Problem 1 were given in [23] and [40] for $D_i^{-1}, i = 1,...,m$, monotone Lipschitzian and cocoercive operators, respectively. Different to the approach proposed in this subsection, there, the operators $D_i^{-1}, i = 1,...,m$, are processed within forward steps.

(ii) When for every $i = 1,...,m$ one takes $D_i(0) = \mathcal{G}_i$ and $D_i(v) = \varnothing\ \forall v \in \mathcal{G}_i \setminus \{0\}$, the algorithms proposed in [23, Theorem 3.1] (see, also, [17, Theorem 3.1] for the case $m = 1$) and [40, Theorem 3.1] applied to Problem 1 differ from Algorithm 11.

(iii) When solving the particular case of a primal-dual pair of convex optimization problems

$$\inf_{x\in\mathcal{H}} \{f(x) + g(Lx)\},$$

and

$$\sup_{v\in\mathcal{G}} \{-f^*(-L^*v) - g^*(v)\},$$

where $f \in \Gamma(\mathcal{H}), g \in \Gamma(\mathcal{G})$ and $L : \mathcal{H} \to \mathcal{G}$ is a linear continuous operator, one can make use of the iterative schemes provided in [24, Algorithm 3.1] and [19, Algorithm 1]. Let us notice that particularizing Algorithm 11 to this framework gives rise to a numerical scheme different to the ones in the mentioned literature.

3.3.2 A second primal-dual Douglas-Rachford type algorithm

In Algorithm 11 each operator L_i and its adjoint $L_i^*, i = 1,...,m$ are processed two times, however, for large-scale optimization problems these matrix-vector multiplications may be expensive compared with the computation of the resolvents of the operators A, B_i^{-1} and $D_i^{-1}, i = 1,...,m$. The second primal-dual algorithm of Douglas-Rachford type we propose for solving the monotone inclusions in Problem 1 has the particularity that it evaluates each operator L_i and its adjoint $L_i^*, i = 1,...,m$, only once.

Algorithm 13. Let $x_0 \in \mathscr{H}$, $(y_{1,0},\dots,y_{m,0}) \in \mathscr{G}_1 \times \dots \times \mathscr{G}_m$, $(v_{1,0},\dots,v_{m,0}) \in \mathscr{G}_1 \times \dots \times \mathscr{G}_m$, and τ and $\sigma_i, i = 1,...,m$, be strictly positive real numbers such that $\tau \sum_{i=1}^m \sigma_i \|L_i\|^2 < \frac{1}{4}$. Furthermore, let $\gamma_i \in \mathbb{R}_{++}$, $\gamma_i \leq 2\sigma_i^{-1}\tau \sum_{i=1}^m \sigma_i \|L_i\|^2, i = 1,...,m$, let $(\lambda_n)_{n\geq 0}$ be a sequence in $(0,2)$ and set

$$(\forall n \geq 0) \quad \left\lfloor \begin{array}{l} p_{1,n} = J_{\tau A}\left(x_n - \tau\left(\sum_{i=1}^m L_i^* v_{i,n} - z\right)\right) \\ x_{n+1} = x_n + \lambda_n(p_{1,n} - x_n) \\ \text{For } i = 1,\dots,m \\ \left\lfloor \begin{array}{l} p_{2,i,n} = J_{\gamma_i D_i}\left(y_{i,n} + \gamma_i v_{i,n}\right) \\ y_{i,n+1} = y_{i,n} + \lambda_n(p_{2,i,n} - y_{i,n}) \\ p_{3,i,n} = J_{\sigma_i B_i^{-1}}\left(v_{i,n} + \sigma_i\left(L_i(2p_{1,n} - x_n) - (2p_{2,i,n} - y_{i,n}) - r_i\right)\right) \\ v_{i,n+1} = v_{i,n} + \lambda_n(p_{3,i,n} - v_{i,n}). \end{array} \right. \end{array} \right. \tag{42}$$

Theorem 14. *In Problem 1 suppose that $Cx = 0$ for all $x \in \mathscr{H}$,*

$$z \in \operatorname{ran}\left(A + \sum_{i=1}^m L_i^*(B_i \Box D_i)(L_i \cdot - r_i)\right) \tag{43}$$

and consider the sequences generated by Algorithm 13.

1. *If $\sum_{n=0}^{+\infty} \lambda_n(2-\lambda_n) = +\infty$, then*
   a. *$(x_n, y_{1,n},\dots,y_{m,n}, v_{1,n},\dots,v_{m,n})_{n\geq 0}$ converges weakly to a point $(\bar{x},\bar{y}_1,...,\bar{y}_m, \bar{v}_1,...,\bar{v}_m) \in \mathscr{H} \times \mathscr{G}_1 \times \dots \times \mathscr{G}_m \times \mathscr{G}_1 \times \dots \times \mathscr{G}_m$ such that $(\bar{x},\bar{v}_1,...,\bar{v}_m)$ is a primal-dual solution to Problem 1.*
   b. *$\lambda_n(p_{1,n} - x_n) \to 0$ $(n \to +\infty)$, $\lambda_n(p_{2,i,n} - y_{i,n}) \to 0$ $(n \to +\infty)$ and $\lambda_n(p_{3,i,n} - v_{i,n}) \to 0$ $(n \to +\infty)$ for $i = 1,...,m$.*
   c. *whenever $\mathscr{H}$ and $\mathscr{G}_i, i = 1,...,m$, are finite-dimensional Hilbert spaces, $(x_n, v_{1,n},\dots,v_{m,n})_{n\geq 0}$ converges strongly to a primal-dual solution of Problem 1.*
2. *If $\inf_{n\geq 0} \lambda_n > 0$ and*

$$A, B_i^{-1} \text{ and } D_i, i = 1,...,m, \text{ are uniformly monotone},$$

then $(p_{1,n}, p_{3,1,n},\dots,p_{3,m,n})_{n\geq 0}$ converges strongly to the unique primal-dual solution of Problem 1.

For the proof of Theorem 14 we refer the reader to [14].

Remark 6. When for every $i = 1,...,m$ one takes $D_i(0) = \mathscr{G}_i$ and $D_i(v) = \varnothing \ \forall v \in \mathscr{G}_i \setminus \{0\}$, and $(d_{i,n})_{n\geq 0}$ as a sequence of zeros, one can show that the assertions made in Theorem 14 hold true for step length parameters satisfying

$$\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < 1,$$

when choosing $(y_{1,0},\dots,y_{m,0}) = (0,\dots,0)$ in Algorithm 13, since the sequences $(y_{1,n},\dots,y_{m,n})_{n\geq 0}$ and $(v_{1,n},\dots,v_{m,n})_{n\geq 0}$ vanish in this particular situation.

Remark 7. In the following we emphasize the relations between Algorithm 13 and other existent primal-dual iterative schemes.

(i) When for every $i = 1,...,m$ one takes $D_i(0) = \mathscr{G}_i$ and $D_i(v) = \varnothing \ \forall v \in \mathscr{G}_i \setminus \{0\}$, Algorithm 13 with $(y_{1,0},\dots,y_{m,0}) = (0,\dots,0)$ as initial choice provides an iterative scheme which is identical to the one in [40, Eq. (3.3)], but differs from the one in [23, Theorem 3.1] (see, also, [17, Theorem 3.1] for the case $m = 1$) when the latter are applied to Problem 1.

(ii) When solving the particular case of a primal-dual pair of convex optimization problems mentioned in Remark 5(iii) and when considering as initial choice $y_{1,0} = 0$, Algorithm 13 gives rise to an iterative scheme which is equivalent to [24, Algorithm 3.1]. Furthermore, the method in Algorithm 13 equals the one in [19, Algorithm 1], our choice of $(\lambda_n)_{n\geq 0}$ to be variable in the interval $(0,2)$, however, relaxes the assumption in [19] that $(\lambda_n)_{n\geq 0}$ is a constant sequence in $(0,1]$.

4 Applications to convex optimization

In this section we will employ theoretical results presented in Section 3 in the context of solving convex optimization problems, an approach which relies on the fruitful idea that the convex subdifferential of a proper, convex and lower semicontinuous function is a maximally monotone operator. We are able to treat a wide variety of real world problems, in fields like image denoising and deblurring, support vector machine learning, location theory, portfolio optimization and clustering, where convex optimization problems solvable via primal-dual splitting algorithms occur. The primal-method for which we opt will be in concordance with the nature of the convex optimization problem under investigation.

4.1 A primal-dual pair of convex optimization problems

The primal-dual pair of convex optimization problems under investigation is described as follows.

Problem 2. For a real Hilbert space $\mathcal{H}$, let $z \in \mathcal{H}$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ be a convex and differentiable function with μ-Lipschitzian gradient with $\mu \in \mathbb{R}_{++}$. Furthermore, for $i = 1, \ldots, m$, consider the real Hilbert space $\mathcal{G}_i$, let $r_i \in \mathcal{G}_i$, $g_i, l_i \in \Gamma(\mathcal{G}_i)$ be such that l_i is ν_i^{-1}-strongly convex with $\nu_i \in \mathbb{R}_{++}$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. We consider the convex minimization problem

$$(P) \quad \inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \Box l_i)(L_i x - r_i) + h(x) - \langle x, z \rangle \right\} \tag{44}$$

and its dual problem

$$(D) \quad \sup_{(v_i, \ldots, v_m) \in \mathcal{G}_1 \times \ldots \times \mathcal{G}_m} \left\{ -(f^* \Box h^*) \left(z - \sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} \left(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i \rangle \right) \right\}. \tag{45}$$

In order to investigate the primal-dual pair (44)-(45) in the context of Problem 1, one has to take

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

Then A and B_i, $i = 1, ..., m$ are maximally monotone, C is monotone and μ^{-1}-cocoercive (resp. μ-Lipschitz continuous), by [2, Proposition 17.10], and $D_i^{-1} = \nabla l_i^*$ is monotone and ν_i^{-1}-cocoercive (resp. ν_i-Lipschitz continuous), $i = 1, \ldots, m$, according to [2, Proposition 17.10, Theorem 18.15 and Corollary 16.24].

Remark 8. When solving monotone inclusion problems arising in convex optimization the problem formulations of the forward-backward and of the forward-backward-forward type algorithms coincide, as a result of the Baillon-Haddad Theorem (cf. [2, Corollary 18.16]). On the other hand, when $h : \mathcal{H} \to \mathbb{R}$, $h(x) = 0$ for all $x \in \mathcal{H}$, one can consider the Douglas-Rachford type methods treated in Section 3.3, even if the strong convexity assumption imposed on the functions $l_i \in \Gamma(\mathcal{G}_i)$, $i = 1, \ldots, m$, are removed.

Whenever $(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \ldots \times \mathcal{G}_m$ is a primal-dual solution to (44)-(45), namely,

$$z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i \overline{x} - r_i), \; i = 1, \ldots, m, \tag{46}$$

then $\overline{x}$ is an optimal solution to (P), $(\overline{v}_1, \ldots, \overline{v}_m)$ is an optimal solution to (D) and the optimal objective values of the two problems, which we denote by $v(P)$ and $v(D)$, respectively, coincide (thus, strong duality holds).

Since a fundamental assumption in the convergence theorems provided in the previous section asks for the existence of a solution for the monotone inclusion problem under investigation, we formulate in the following proposition, which was given in [23, Proposition 4.3], sufficient conditions for it in the context of convex optimization problems. To this end we mention that the *strong quasi-relative interior* of a nonempty convex set $\Omega \subseteq \mathscr{H}$ is defined as

$$\operatorname{sqri}\Omega = \left\{ x \in \Omega : \bigcup_{\lambda \geq 0} \lambda(\Omega - x) \text{ is a closed linear subspace} \right\}.$$

Proposition 1. *Suppose that* (P) *has at least one solution and set*

$$S := \{(L_1 x - y_1, ..., L_m x - y_m) : x \in \operatorname{dom} f \text{ and } y_i \in \operatorname{dom} g_i + \operatorname{dom} l_i, i = 1, ..., m\}.$$

The inclusion

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

is satisfied, if one of the following holds:

(i) $(r_1, \dots, r_m) \in \operatorname{sqri} S$.
(ii) for every $i \in \{1, \dots, m\}$, g_i *or* l_i *is real-valued.*
(iii) $\mathscr{H}$ *and* $\mathscr{G}_i, i = 1, ..., m$, *are finite dimensional and there exists* $x \in \operatorname{ri}\operatorname{dom} f$ *such that*

$$L_i x - r_i \in \operatorname{ri}\operatorname{dom} g_i + \operatorname{ri}\operatorname{dom} l_i, \; i = 1, ..., m.$$

4.2 Image processing involving total variation functionals

For the applications discussed in the context of image processing, the images have been normalized, in order to make their pixels range in the closed interval from 0 to 1.

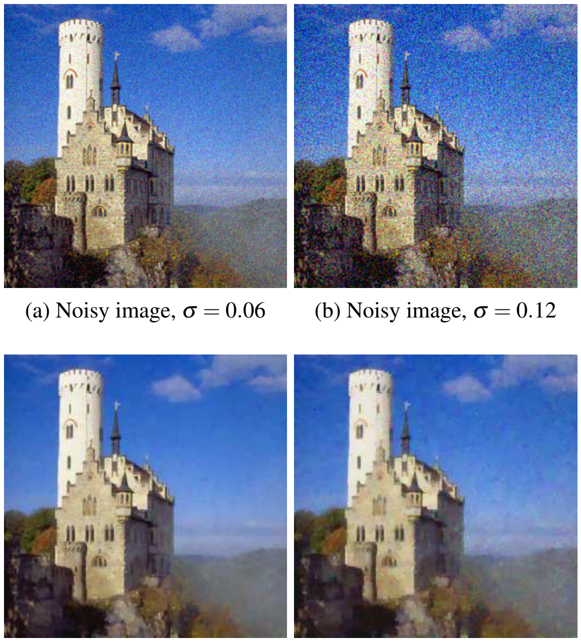
4.2.1 TV-based image denoising

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

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

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 $j = 1, \dots, N$. Two popular choices for the



(a) Noisy image, $\sigma = 0.06$ (b) Noisy image, $\sigma = 0.12$

(c) Denoised image, $\lambda = 0.035$ (d) Denoised image, $\lambda = 0.07$

Fig. 1 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 (c) shows the denoised image for $\lambda = 0.035$. Likewise, the noisy image when choosing $\sigma = 0.12$ and the denoised one for $\lambda = 0.07$ are shown in (b) and (d), respectively.

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.

We denote $\mathscr{Y} = \mathbb{R}^n \times \mathbb{R}^n$ and define the linear operator $L : \mathbb{R}^n \to \mathscr{Y}$, $x_{i,j} \mapsto (L_1 x_{i,j}, L_2 x_{i,j})$, where

$$L_1 x_{i,j} = \begin{cases} x_{i+1,j} - x_{i,j}, & \text{if } i < M \\ 0, & \text{if } i = M \end{cases} \text{ and } L_2 x_{i,j} = \begin{cases} x_{i,j+1} - x_{i,j}, & \text{if } j < N \\ 0, & \text{if } j = N \end{cases}.$$

	$\sigma = 0.12$, $\lambda = 0.07$		$\sigma = 0.06$, $\lambda = 0.035$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$
DR1	1.14s (48)	2.80s (118)	1.07s (45)	2.44s (103)
DR2	0.92s (75)	2.10s (173)	0.80s (66)	1.78s (147)
FBF	7.51s (343)	49.66s (2271)	4.08s (187)	34.44s (1586)
FBF Acc	2.20s (101)	9.84s (451)	1.61s (73)	6.70s (308)
PD	3.69s (337)	24.34s (2226)	2.02s (183)	16.74s (1532)
PD Acc	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)
AMA Acc	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 1: Performance evaluation for the images in Figure 1. The entries refer to the CPU times in seconds and to the number of iterations, respectively, needed in order to attain a root mean squared error for the iterates below the tolerance ε.

The operator L represents a discretization of the gradient using reflexive (Neumann) boundary conditions and standard finite differences. One can easily check that $\|L\|^2 \leq 8$, while its adjoint $L^* : \mathscr{Y} \to \mathbb{R}^n$ is given in [18].

Within this example we will focus on the anisotropic total variation functional, which is nothing else than the composition of the l_1-norm on $\mathscr{Y}$ with the linear operator L. Due to the full splitting characteristics of the iterative methods presented in the previous sections, we only need to compute the proximal point of the conjugate of the l_1-norm, the latter being the indicator function of the dual unit ball. Thus, the calculation of the proximal point will result in the computation of a projection, which admits an efficient implementation. The more challenging isotropic total variation functional is employed in the forthcoming subsection in the context of image deblurring.

Thus, problem (47) reads equivalently

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

where $h : \mathbb{R}^n \to \mathbb{R}$, $h(x) = \frac{1}{2}\|x - b\|^2$, is 1-strongly convex and differentiable with 1-Lipschitzian gradient and $g : \mathscr{Y} \to \mathbb{R}$ is defined as $g(y_1, y_2) = \lambda\|(y_1, y_2)\|_1$. Then its conjugate $g^* : \mathscr{Y} \to \overline{\mathbb{R}}$ 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 the two Douglas-Rachford type primal-dual methods (DR1, cf. Algorithm 11, and DR2, cf. Algorithm 13), the forward-backward-forward type primal dual method (FBF, cf. Theorem 6) and its accelerated version (FBF Acc, cf. Algorithm 7), the primal-dual method (PD) and its accelerated version (PD Acc), both given in [19], the alternating minimization algorithm (AMA) from [38] together with its

Nesterov-type acceleration (cf. [33]), as well as the Nesterov (cf. [32]) and FISTA (cf. [3]) algorithm operating on the dual problem. A comparison of the obtained results is shown in Table 1.

4.2.2 TV-based image deblurring

The second numerical experiment in image processing concerns the solving of an ill-conditioned linear inverse problem arising in image deblurring. For a given matrix $A \in \mathbb{R}^{n \times n}$ describing a blur (or averaging) operator and a given vector $b \in \mathbb{R}^n$ representing the blurred and noisy image, our aim is to estimate the unknown original image $\overline{x} \in \mathbb{R}^n$ fulfilling

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

To this end we solved the following regularized convex nondifferentiable problem

$$\inf_{x \in \mathbb{R}^n} \Big\{ \|Ax - b\|_1 + \alpha_2 \|Wx\|_1 + \alpha_1 TV(x) + \delta_{[0,1]^n}(x) \Big\}, \tag{48}$$

where the regularization is done by a combination of two functionals with different properties. Here, $\alpha_1, \alpha_2 \in \mathbb{R}_{++}$ are regularization parameters, $TV : \mathbb{R}^n \to \mathbb{R}$ is the discrete isotropic total variation function and $W : \mathbb{R}^n \to \mathbb{R}^n$ is the discrete Haar wavelet transform with four levels.

For $(y,z), (p,q) \in \mathscr{Y}$, 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 $\mathscr{Y}$ and that for every $x \in \mathbb{R}^n$ it holds $TV_{\mathrm{iso}}(x) = \|Lx\|_\times$, where L is the linear operator defined in the previous subsection.

Consequently, the optimization problem (48) can be equivalently written as

$$\inf_{x \in \mathbb{R}^n} \{ f(x) + g_1(Ax) + g_2(Wx) + g_3(Lx) \}, \tag{49}$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $f(x) = \delta_{[0,1]^n}(x)$, $g_1 : \mathbb{R}^n \to \mathbb{R}$, $g_1(y) = \|y - b\|_1$, $g_2 : \mathbb{R}^n \to \mathbb{R}$, $g_2(y) = \alpha_2 \|y\|_1$ and $g_3 : \mathscr{Y} \to \mathbb{R}$, $g_3(y,z) = \alpha_1 \|(y,z)\|_\times$. The proximal points of these functions admit explicit representations (see, for instance, [13, 14]).

Figure 2 shows the performance of Algorithm 11 (DR1) and Algorithm 13 (DR2) when solving (49) for $\alpha_1 = 3\text{e-}3$ and $\alpha_2 = 1\text{e-}3$. It also shows the original, observed and reconstructed versions of the 256×256 cameraman test image.

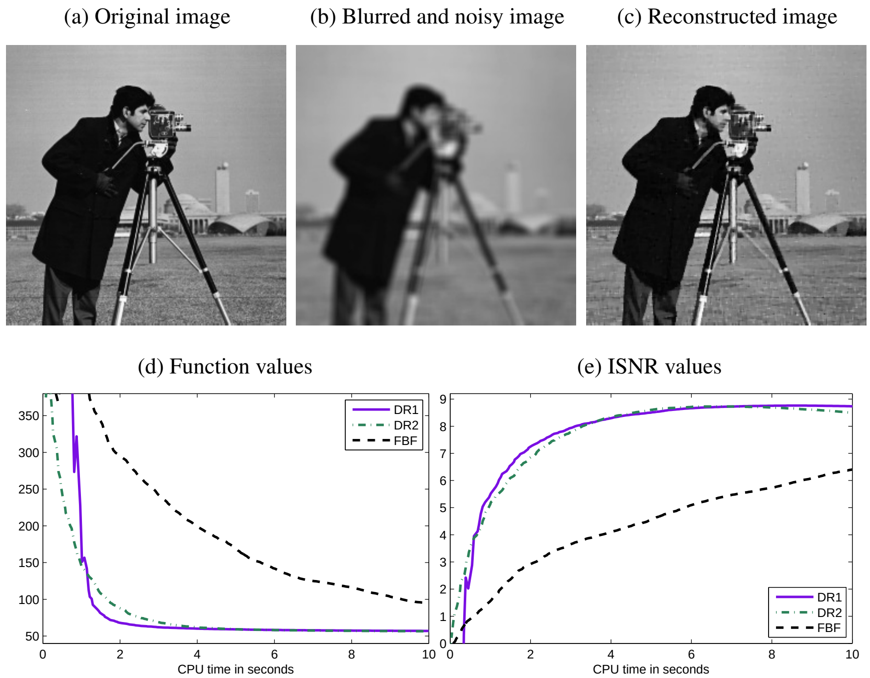


Fig. 2: Top: Original, observed and reconstructed versions of the cameraman. Bottom: The evolution of the values of the objective function and of the ISNR (improvement in signal-to-noise ratio) for Algorithm 11 (DR1), Algorithm 13 (DR2) and the forward-backward-forward method (FBF) from Theorem 6.

4.3 Kernel based machine learning

The next numerical experiment concerns the solving of the problem of classifying images via support vector machines classification, an approach which belongs to the class of kernel based learning methods.

The given data set consisting of 11339 training images and 1850 test images of size 28×28 was taken from the website http://www.cs.nyu.edu/ roweis/data.html. The problem we consider is to determine a decision function based on a pool of handwritten digits showing either the number five or the number six, labeled by $+1$ and -1, respectively (see Figure 3). Subsequently, we evaluate the quality of the decision function on the test data set by computing the percentage of misclassified images. In order to reduce the computational effort, we used only half of the available images from the training data set.

The classifier functional $\mathtt{f}$ is assumed to be an element of the *Reproducing Kernel Hilbert Space (RHKS)* $\mathscr{H}_\kappa$, which in our case is induced by the symmetric and finitely positive definite Gaussian kernel function

Fig. 3: A sample of images belonging to the classes $+1$ and -1, respectively.

$$\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R},\ \kappa(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma_\kappa^2}\right).$$

Let $\langle\cdot,\cdot\rangle_\kappa$ denote the inner product on $\mathscr{H}_\kappa$, $\|\cdot\|_\kappa$ the corresponding norm and $K \in \mathbb{R}^{n\times n}$ the *Gram matrix* with respect to the training data set

$$\mathscr{Z} = \{(X_1,Y_1),\dots,(X_n,Y_n)\} \subseteq \mathbb{R}^d \times \{+1,-1\},$$

namely the symmetric and positive definite matrix with entries $K_{ij} = \kappa(X_i,X_j)$ for $i,j = 1,\dots,n$. Within this example we make use of the *hinge loss* $v : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, $v(x,y) = \max\{1-xy,0\}$, which penalizes the deviation between the predicted value $\mathtt{f}(x)$ and the true value $y \in \{+1,-1\}$. The smoothness of the decision function $\mathtt{f} \in \mathscr{H}_\kappa$ is employed by means of the *smoothness functional* $\Omega : \mathscr{H}_\kappa \to \mathbb{R}$, $\Omega(f) = \|\mathtt{f}\|_\kappa^2$, taking high values for non-smooth functions and low values for smooth ones. The decision function $\mathtt{f}$ we are looking for is the optimal solution of the *Tikhonov regularization problem*

$$\inf_{\mathtt{f}\in\mathscr{H}_\kappa} \left\{ C\sum_{i=1}^n v(\mathtt{f}(X_i),Y_i) + \frac{1}{2}\Omega(\mathtt{f}) \right\}, \tag{50}$$

where $C > 0$ denotes the regularization parameter controlling the tradeoff between the loss function and the smoothness functional.

The *representer theorem* (cf. [36]) ensures the existence of a vector of coefficients $c = (c_1,\dots,c_n)^T \in \mathbb{R}^n$ such that the minimizer $\mathtt{f}$ of (50) can be expressed as a kernel expansion in terms of the training data, i.e., $\mathtt{f}(\cdot) = \sum_{i=1}^n c_i\kappa(\cdot,X_i)$. Thus, the smoothness functional becomes $\Omega(\mathtt{f}) = \|\mathtt{f}\|_\kappa^2 = \langle \mathtt{f},\mathtt{f}\rangle_\kappa = \sum_{i=1}^n\sum_{j=1}^n c_ic_j\kappa(X_i,X_j) = c^TKc$ and for $i = 1,\dots,n$ it holds $\mathtt{f}(X_i) = \sum_{j=1}^n c_j\kappa(X_i,X_j) = (Kc)_i$. Hence, in order to determine the decision function one has to solve the convex optimization problem

$$\inf_{c\in\mathbb{R}^n} \{g(Kc) + h(c)\}, \tag{51}$$

where $g : \mathbb{R}^n \to \mathbb{R}$, $g(z) = C\sum_{i=1}^n v(z_i,Y_i)$, and $h : \mathbb{R}^n \to \mathbb{R}$, $h(c) = \frac{1}{2}c^TKc$. The function $h : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable and it fulfills $\nabla h(c) = Kc$ for every $c \in \mathbb{R}^n$, thus ∇h is Lipschitz continuous with constant $\mu = \|K\|$. It is much easier to process the function h via its gradient than via its proximal point. For every $p \in \mathbb{R}^n$ it holds (see, also, [12, 16])

$$g^*(p) = \sup_{z\in\mathbb{R}^n}\left\{\langle p,z\rangle - C\sum_{i=1}^n v(z_i,Y_i)\right\} = \sum_{i=1}^n (Cv(\cdot,Y_i))^*(p_i) = C\sum_{i=1}^n v(\cdot,Y_i)^*\left(\frac{p_i}{C}\right)$$
$$= \begin{cases} \sum_{i=1}^n p_iY_i, & \text{if } p_iY_i \in [-C,0],\ i=1,\dots,n,\\ +\infty, & \text{otherwise.}\end{cases}$$

Thus, for $\sigma \in \mathbb{R}_{++}$ and $c \in \mathbb{R}^n$ we have

$$\begin{aligned}\operatorname{Prox}_{\sigma g^*}(c) &= \underset{p\in\mathbb{R}^n}{\arg\min}\left\{\sigma C\sum_{i=1}^n v(\cdot,Y_i)^*\left(\frac{p_i}{C}\right) + \frac{1}{2}\|p-c\|^2\right\}\\ &= \underset{\substack{p_iY_i\in[-C,0]\\ i=1,\dots,n}}{\arg\min}\left\{\sum_{i=1}^n\left[\sigma p_iY_i + \frac{1}{2}(p_i-c_i)^2\right]\right\}\\ &= \left(\mathscr{P}_{Y_1[-C,0]}(c_1-\sigma Y_1),\dots,\mathscr{P}_{Y_n[-C,0]}(c_n-\sigma Y_n)\right)^T.\end{aligned}$$

With respect to the considered dataset, we denote by

$$\mathscr{D} = \{(X_i,Y_i), i=1,\dots,5670\} \subseteq \mathbb{R}^{784}\times\{+1,-1\}$$

the set of available training data consisting of 2711 images in the class $+1$ and 2959 images in the class -1. Notice that a sample from each class of images is shown in Figure 3. Due to numerical reasons, the images have been normalized (cf. [28]) by dividing each of them by the quantity $\left(\frac{1}{5670}\sum_{i=1}^{5670}\|X_i\|^2\right)^{\frac{1}{2}}$.

C	kernel parameter σ_κ					
	0.125	0.25	0.5	0.75	1	2
0.1	1.0270	1.3514	1.3514	1.8919	2.1081	3.0270
1	1.0270	0.7027	0.7568	1.3514	1.4595	2.2162
10	1.0270	0.7568	0.9189	1.0811	1.1892	1.8378
100	1.0270	0.7568	0.8649	1.4054	1.2432	1.8378
1000	1.0270	0.7568	0.8649	1.4595	1.2432	1.8378

Table 2: Misclassification rate in percentage for different model parameters.

In order to determine a good choice for the kernel parameter $\sigma_\kappa \in \mathbb{R}_{++}$ and the tradeoff parameter $C \in \mathbb{R}_{++}$, we tested different combinations of them with the forward-backward (FB) solver given in [40]. The results are shown in Table 2, whereby the combination $\sigma = 0.25$ and $C = 1$ provides with 0.7027 % the lowest misclassification rate. This means that among the 1870 images belonging to the test data set, 13 of them were not correctly classified.

Table 3 shows some results when solving the classification problem (51) via those primal-dual splitting methods which are able to perform a forward step on the operator ∇h. Since the matrix $K \in \mathbb{R}^{n\times n}$ is positive definite, the function

	misclassification rate at 0.7027 %	RMSE$\leq 10^{-3}$
FB	3.07s (113)	19.50s (717)
FB Acc	95.33s (3522)	348.41s (12923)
FBF	4.36s (80)	32.92s (606)
FBF Acc	3.63s (67)	32.90s (606)

Table 3: Performance evaluation for the SVM problem for $C = 1$ and $\sigma_\kappa = 0.25$. The entries refer to the CPU times in seconds and the number of iterations.

$h : \mathbb{R}^n \to \mathbb{R}, h(c) = \frac{1}{2}c^T Kc$, is strongly convex, as well. Hence, there exists a unique solution to (51) and we can also apply the accelerated versions of the (FB) and of the (FBF) method described in Section 3. However, we notice that the acceleration of the forward-backward primal-dual method (FB Acc) converges extremely slow for this instance.

4.4 The generalized Heron problem

The following numerical experiments address the *generalized Heron problem* which has been recently investigated in [30, 31] and where for its solving subgradient-type methods have been used.

While the *classical Heron problem* concerns the finding of a point $\overline{u}$ on a given straight line in the plane such that the sum of distances from $\overline{u}$ to given points u^1, u^2 is minimal, the problem that we address here aims to find a point in a nonempty convex closed set $\Omega \subseteq \mathbb{R}^n$ which minimizes the sum of the distances to given convex closed sets $\Omega_i \subseteq \mathbb{R}^n$, $i = 1,\dots,m$.

The distance from a point $x \in \mathbb{R}^n$ to a nonempty set $\Omega \subseteq \mathbb{R}^n$ is given by

$$d(x;\Omega) = (\|\cdot\| \,\Box\, \delta_\Omega)(x) = \inf_{z\in\Omega} \|x - z\|.$$

Thus the *generalized Heron problem* we address as follows reads

$$\inf_{x\in\Omega} \sum_{i=1}^{m} d(x;\Omega_i). \tag{52}$$

We observe that, due to the formulation of the distance function as the infimal convolution of two proper, convex and lower semicontinuous functions, (52) perfectly fits into the framework considered in Problem 2 and for which the Douglas-Rachford type algorithms were proposed, when setting

$$f = \delta_\Omega, \text{ and } g_i = \|\cdot\|,\ l_i = \delta_{\Omega_i},\ i = 1,\dots,m. \tag{53}$$

However, note that (52) can be solved neither via the forward-backward type nor the forward-backward-forward type primal-dual method, since both of them require the presence of at least one strongly convex function (cf. Baillon-Haddad Theorem, [2, Corollary 18.16]) in each of the infimal convolutions $\|\cdot\| \Box \delta_{\Omega_i}$, $i = 1,\ldots,m$, fact which is obviously here not the case. Notice that

$$g_i^* : \mathbb{R}^n \to \overline{\mathbb{R}},\ g_i^*(p) = \sup_{x\in\mathbb{R}^n} \{\langle p,x\rangle - \|x\|\} = \delta_{B(0,1)}(p),\ i = 1,...,m,$$

thus the proximal points of f, g_i^* and l_i^*, $i = 1,...,m$, can be calculated via projections, in case of the latter via Moreau's decomposition formula.

In the following we test our algorithms on some examples taken from [30, 31].

Example 1 (Example 5.5 in [31]). Consider problem (52) with the constraint set Ω being the closed ball centered at $(5,0)$ having radius 2 and the sets Ω_i, $i = 1,\ldots,8$, being pairwise disjoint squares in *right position* in $\mathbb{R}^2$ (i.e. the edges are parallel to the x- and y-axes, respectively), with centers $(-2,4)$, $(-1,-8)$, $(0,0)$, $(0,6)$, $(5,-6)$, $(8,-8)$, $(8,9)$ and $(9,-5)$ and side length 1, respectively (see Figure 4 (a)).

Figure 4 gives an insight into the performance of the proposed primal-dual methods when compared with the subgradient algorithm used in [31]. After a few milliseconds both splitting algorithms reach machine precision with respect to the *root-mean-square error* where the following parameters were used:

- DR1: $\sigma_i = 0.15$, $\tau = 2/(\sum_{j=1}^{8}\sigma_j)$, $\lambda_n = 1.5$, $x_0 = (5,2)$, $v_{i,0} = 0$, $i = 1,...,8$;
- DR2: $\sigma_i = 0.1$, $\tau = 0.24/(\sum_{j=1}^{8}\sigma_j)$, $\lambda_n = 1.8$, $x_0 = (5,2)$, $v_{i,0} = 0$, $i = 1,...,8$;
- Subgradient (cf. [31, Theorem 4.1]) $x_0 = (5,2)$, $\alpha_n = \frac{1}{n}$.

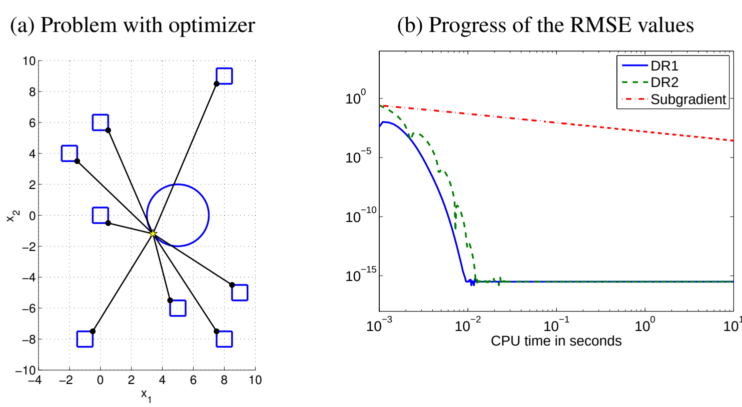


Fig. 4: Example 1. Generalized Heron problem with squares and disc constraint set on the left-hand side, performance evaluation for the *root-mean-square error* (RMSE) on the right-hand side.

Example 2 (Example 4.3 in [30]). In this example we solve the generalized Heron problem (52) in $\mathbb{R}^3$, where the constraint set Ω is the closed ball centered at $(0,2,0)$

with radius 1 and Ω_i, $i = 1,...,5$, are cubes in right position with center at $(0,-4,0)$, $(-4,2,-3)$, $(-3,-4,2)$, $(-5,4,4)$ and $(-1,8,1)$ and side length 2, respectively.

Figure 5 shows that also for this instance the primal-dual approaches outperform the subgradient method from [31]. In this example we used the following parameters:

- DR1: $\sigma_i = 0.3$, $\tau = 2/(\sum_{j=1}^5 \sigma_j)$, $\lambda_n = 1.5$, $x_0 = (0,2,0)$, $v_{i,0} = 0$, $i = 1,...,5$;
- DR2: $\sigma_i = 0.2$, $\tau = 0.24/(\sum_{j=1}^5 \sigma_j)$, $\lambda_n = 1.8$, $x_0 = (0,2,0)$, $v_{i,0} = 0$, $i = 1,...,5$;
- Subgradient (cf. [30, Theorem 4.1]) $x_0 = (5,2)$, $\alpha_n = \frac{1}{n}$.

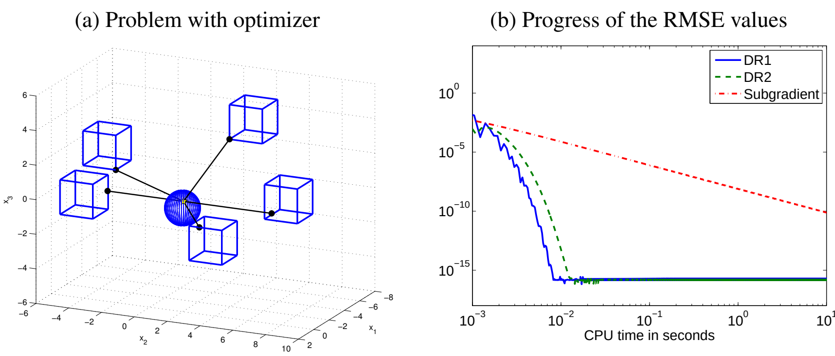


Fig. 5: Example 2. Generalized Heron problem with cubes and ball constraint set on the left-hand side, performance evaluation for the RMSE on the right-hand side.

4.5 Portfolio optimization under different risk measures

We let $(\Omega, \mathfrak{F}, \mathbb{P})$ be an atomless probability space, where the elements ω of Ω represent future states, or individual scenarios (and are allowed to be only finitely many), $\mathfrak{F}$ is a *σ-algebra* on measurable subsets of Ω and $\mathbb{P}$ is a *probability measure* on $\mathfrak{F}$. For a measurable random variable $X : \Omega \to \mathbb{R} \cup \{+\infty\}$ the *expectation value* with respect to $\mathbb{P}$ is defined by $\mathbb{E}[X] := \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$.

Consider further the real Hilbert space

$$L^2 := \left\{ X : \Omega \to \mathbb{R} \cup \{+\infty\} : X \text{ is measurable, } \int_\Omega |X(\omega)|^2\,\mathrm{d}\mathbb{P}(\omega) < +\infty \right\}$$

endowed with *inner product* and *norm* defined for arbitrary $X, Y \in L^2$ via

$$\langle X, Y \rangle = \int_\Omega X(\omega) Y(\omega)\,\mathrm{d}\mathbb{P}(\omega) \text{ and } \|X\| = (\langle X, X \rangle)^{\frac{1}{2}} = \left(\int_\Omega (X(\omega))^2\,\mathrm{d}\mathbb{P}(\omega) \right)^{\frac{1}{2}},$$

respectively.

In this section we measure risk with the so-called Optimized Certainty Equivalent (OCE), which was introduced for concave utility functions in [4, 5] and adapted to convex utility functions in [10]. Here, we call $u : \mathbb{R} \to \overline{\mathbb{R}}$ *utility function*, when u is proper, convex, lower semicontinuous and nonincreasing function such that $u(0) = 0$ and $-1 \in \partial u(0)$.

The generalized convex risk measure we use in order to quantify the risk is defined as (cf. [4, 5, 10])

$$\rho_u : L^2 \to \mathbb{R} \cup \{+\infty\},\ \rho_u(X) = \inf_{\lambda \in \mathbb{R}} \{\lambda + \mathbb{E}[u(X+\lambda)]\}. \tag{54}$$

We consider a portfolio with a number of $N \geq 1$ different positions with returns $R_i \in L^2$, $i = 1, \ldots, N$, a nonzero vector of expected returns $\mu = (\mathbb{E}[R_1], \ldots, \mathbb{E}[R_N])^T$ and $\mu^* \leq \max_{i=1,\ldots,N} \mathbb{E}[R_i]$ a given lower bound for the expected return of the portfolio. In the following, by making use of different utility functions, we are solving the optimization problem

$$\inf_{\substack{x^T\mu \geq \mu^*,\ x^T \mathbb{1}^N = 1, \\ x = (x_1,\ldots,x_N)^T \in \mathbb{R}^N_+}} \rho_u \left(\sum_{i=1}^N x_i R_i \right), \tag{55}$$

which assumes the minimization of the risk of the portfolio subject to constraints on the expected return of the portfolio and on the budget. Here, $\mathbb{1}^N$ denotes the vector in $\mathbb{R}^N$ having all entries equal to 1. By using (54), we obtain the following reformulation of problem (55)

$$\inf_{\substack{x^T\mu \geq \mu^*,\ x^T \mathbb{1}^N = 1, \\ x = (x_1,\ldots,x_N)^T \in \mathbb{R}^N_+,\ \lambda \in \mathbb{R}}} \left\{ \lambda + \mathbb{E}\left[u \left(\sum_{i=1}^N x_i R_i + \lambda \right) \right] \right\}, \tag{56}$$

which will prove to be more suitable for being solved by means of primal-dual proximal splitting algorithms. Therefore, we introduce the convex closed sets

$$S = \left\{ x \in \mathbb{R}^N : x^T \mu \geq \mu^* \right\} \text{ and } T = \left\{ x \in \mathbb{R}^N : x^T \mathbb{1}^N = 1 \right\},$$

and reformulate (56) as the unconstrained problem

$$\inf_{(x,\lambda) \in \mathbb{R}^N \times \mathbb{R}} \left\{ \delta_{\mathbb{R}^N_+}(x) + \lambda + \delta_{S \times \mathbb{R}}(x,\lambda) + \delta_{T \times \mathbb{R}}(x,\lambda) + (\mathbb{E}[u] \circ K)(x,\lambda) \right\}, \tag{57}$$

where $K : \mathbb{R}^N \times \mathbb{R} \to L^2$, $(x_1, \ldots, x_n, \lambda) \mapsto \sum_{i=1}^N x_i R_i + \lambda$. When calculating the proximal points of the functions occurring in the formulation of this convex minimization problem, one has only to determine the projections on the sets $\mathbb{R}^N_+$, S, and T, for which explicit formulae can be given (cf. [2, Example 3.21 and Example 28.16]). The proximal point with respect to the function $\mathbb{E}[u]$ can be obtained via the following proposition (cf. [15]).

μ^*	linear ($\alpha = 0.95$)	exponential	indicator	quadr. ($\beta = 1$)	log. ($\theta = 5$)
0.3	0.14s (500)	0.18s (402)	- (> 15000)	0.05s (170)	0.53s (1891)
0.5	0.15s (520)	0.15s (336)	- (> 15000)	0.06s (196)	0.38s (1335)
0.7	0.33s (1202)	0.31s (682)	- (> 15000)	0.06s (186)	0.72s (2570)
0.9	0.32s (1164)	0.40s (885)	- (> 15000)	0.08s (272)	1.07s (3820)
1.1	0.41s (1526)	6.80s (15222)	- (> 15000)	0.14s (486)	1.18s (4198)
1.3	0.42s (1570)	5.45s (12155)	- (> 15000)	0.41s (1476)	6.61s (23547)

Table 4: CPU times in seconds and the number of iterations when solving the portfolio optimization problem (55) for different utility functions.

Proposition 2. *For arbitrary random variables $X \in L^2$ and $\gamma \in \mathbb{R}_{++}$ it holds*

$$\mathrm{Prox}_{\gamma \mathbb{E}[u]}(X)(\omega) = \mathrm{Prox}_{\gamma u}(X(\omega)) \ \forall \omega \in \Omega \ a.\,s.. \tag{58}$$

For our experiments we took weekly opening courses over the last 13 years from assets belonging to the indices DAX and NASDAQ in order to obtain the returns $R_i \in \mathbb{R}^{|\Omega|}, i = 1,\ldots,N$, for $|\Omega| = 689$ and $N = 106$. The data was provided by the Yahoo finance database. Assets which do not support the required historical information like Volkswagen AG (DAX) or Netflix, Inc. (NASDAQ) were not taken into account.

For solving the portfolio optimization problem (55), we used convex risk measures induced by linear, exponential, indicator, quadratic and logarithmic utility functions. We applied Algorithm 11 (DR1) for solving the unconstrained problem in (57), while using formulae for the proximal points of each utility function given in [15]. The values of the expected returns associated with R_i, $i = 1,\ldots,N$ ranged from -0.2690 (Commerzbank AG, DAX) to 1.4156 (priceline.com Incorporated, NASDAQ).

Computational results for this problem are reported in Table 4 for different values of μ^*. We terminated the algorithm when subsequent iterates started to stay within an accuracy level of 1% with respect to the set of constraints and to the optimal objective value. It shows that the worst-case risk measure, which is obtained by using the indicator utility, performs poorly on the given dataset, while it seems that the algorithm is sensitive to the lower bound on the expected return μ^*.

4.6 Clustering

In cluster analysis one aims for grouping a set of points such that points within the same group are more similar (usually measured via distance functions) to each other than to points in other groups. Clustering can be formulated as a convex optimization problem (see, for instance, [27, 29, 20]). In this example, we are treating the minimization problem

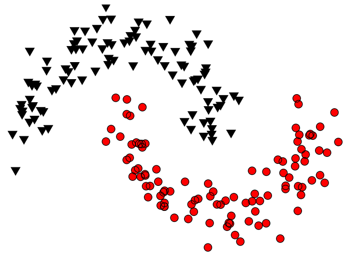


Fig. 6 Clustering two interlocking half moons. The colors (resp. the shapes) show the correct affiliations.

$$\inf_{x_i \in \mathbb{R}^n, i=1,\dots,m} \left\{ \frac{1}{2} \sum_{i=1}^{m} \|x_i - u_i\|^2 + \gamma \sum_{i<j} \omega_{ij} \|x_i - x_j\|_p \right\}, \tag{59}$$

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. In [27], the authors consider ℓ_1, ℓ_2, and ℓ_∞ norms on the penalty terms $x_i - x_j$ while in [29] arbitrary ℓ_p norms were taken into account. Since the objective function is strongly convex, there exists a unique solution to (59).

The tuning parameter $\gamma \in \mathbb{R}_+$ plays a central role in 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 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 promptly as γ passes certain critical values. For our weights we used a K-nearest neighbors strategy, as proposed in [20]. 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 took the values $K = 10$ and $\phi = 0.5$, which are the best ones reported in [20] on a similar dataset.

Due to the nature of the optimization problem under investigation, all primal-dual splitting methods presented in this paper could be employed in order to solve it. Independently on the choice of $p \in \{1,2\}$ we took profit in our implementations from the exact representations of the proximal points of all functions involved in its formulation.

For our numerical tests we considered the standard dataset consisting of two interlocking half moons in $\mathbb{R}^2$, each of them being composed of 100 points (see Figure 6). After determining the unique optimal solution with a state-of-the-art solver over 100000 iterations, our stopping criterion asks the *root mean squared error* (RMSE) to be less than or equal to a given bound ε. As tuning parameters we used $\gamma = 4$ for

	$p=2, \gamma=5.2$		$p=1, \gamma=4$	
	$\varepsilon=10^{-4}$	$\varepsilon=10^{-8}$	$\varepsilon=10^{-4}$	$\varepsilon=10^{-8}$
DR1	0.78s (216)	1.68s (460)	0.78s (218)	1.68s (464)
DR2	0.61s (323)	1.20s (644)	0.60s (325)	1.18s (648)
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)
FB	2.48s (1353)	5.72s (3090)	2.01s (1092)	4.05s (2226)
FB Acc	2.04s (1102)	4.11s (2205)	1.74s (950)	3.84s (2005)
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	7.85s (3811)	42.69s (21805)	7.46s (3936)	$>$ 190s ($>$ 100000)
FISTA	7.55s (4055)	51.01s (27356)	6.55s (3550)	47.81s (26069)

Table 5: 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.

$p=1$ and $\gamma=5.2$ for $p=2$. 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 5, it shows that the two Douglas-Rachford type primal-dual methods are superior to all other algorithms within this comparison. One can also see that the accelerations of the forward-backward-forward (FBF) and of the forward-backward (FB) type primal-dual methods have a positive effect on both CPU times and required iterations compared with the regular methods. The alternating minimization algorithm (AMA, cf. [38]) converges slow in this example. Its Nesterov-type acceleration (cf. [33]), however, performs better. The two accelerated first-order methods FISTA (cf. [3]) and the one we called Nesterov (cf. [32]), which are both solving the dual problem, perform surprisingly bad in this case. In the numerical experiments example on image denoising both methods proved to have a good performance.

References

1. Attouch, H., Théra, M.: A general duality principle for the sum of two operators. J. Convex Anal. **3**(1), 1–24 (1996)
2. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. CMS Books in Mathematics, Springer, New York (2011)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-tresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
4. Ben-Tal, A., Teboulle, M.: Expected utility, penalty functions and duality in stochastic non-linear programming. Manag. Sci. **32**(11), 1445–1466 (1986)
5. Ben-Tal, A., Teboulle, M.: An old-new concept of risk measures: the optimized certainty equivalent. Math. Finance **17**(3), 449–476 (2007)
6. Borwein, J.M., Vanderwerff, J.D.: Convex Functions: Constructions, Characterizations and Counterexamples. Cambridge University Press, Cambridge (2010)

7. Boţ, R.I.: Conjugate Duality in Convex Optimization. Lecture Notes in Economics and Mathematical Systems, Vol. 637, Springer, Berlin (2010)
8. Boţ, R.I., Csetnek, E.R., Heinrich, A.: A primal-dual splitting algorithm for fnding zeros of sums of maximally monotone operators. arXiv:1206.5953 [math.OC]. To appear in SIAM J. Optim. (2012)
9. Boţ, R.I., Csetnek, E.R., Heinrich, A.: On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems. arXiv:1303.2875 [math.OC] (2013)
10. Boţ, R.I., Frătean, A.R.: Looking for appropriate qualification conditions for subdifferential formulae and dual representations for convex risk measures. Math. Meth. Oper. Res. **74**(2), 191–215 (2011)
11. Boţ, R.I., Grad, S.M., Wanka,G.: Duality in Vector Optimization. Springer, Berlin (2009)
12. Boţ, R.I., Heinrich, A., Wanka, G.: Employing different loss functions for the classification of images via supervised learning. Preprint, Chemnitz University of Technology, Faculty of Mathematics (2012)
13. Boţ, R.I., Hendrich, C.: Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization. arXiv:1211.1706v1 [math.OC] (2012)
14. Boţ, R.I., Hendrich, C.: A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators. arXiv:1212.0326 [math.OC] (2012)
15. Boţ, R.I., Hendrich, C.: Convex risk minimization via proximal splitting methods. arXiv: 1304.7694 [math.OC] (2013)
16. Boţ, R.I., Lorenz, N.: Optimization problems in statistical learning: Duality and optimality conditions. Eur. J. Oper. Res. **213**(2), 395–404 (2011)
17. Briceño-Arias, L.M., Combettes, P.L.: A monotone + skew splitting model for composite monotone inclusions in duality. SIAM J. Optim. **21**(4), 1230–1250 (2011)
18. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**(1–2), 89–97 (2004)
19. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)
20. Chi, E.C., Lange, K.: Splitting methods for convex clustering. arXiv:1304.0499 [stat.ML] (2013)
21. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**(5–6), 475–504 (2004)
22. Combettes, P.L.: Iterative construction of the resolvent of a sum of maximal monotone operators. J. Convex Anal. **16**(3), 727–748 (2009)
23. Combettes, P.L., Pesquet, J.-C.: 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)
24. Condat., L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**(2), 460–479 (2013)
25. Douglas, J., Rachford, H.H.: On the numerical solution of the heat conduction problem in 2 and 3 space variables. Trans. Amer. Math. Soc **82**(2), 421–439 (1956)
26. Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. North-Holland Publishing Company, Amsterdam (1976)
27. Hocking, T., Vert, J., Bach, F., Joulin, A.: Clusterpath: an algorithm for clustering using convex fusion penalties. In: ICML (2011)
28. Lal, T.N., Chapelle, O., Schölkopf, B.: Combining a filter method with SVMs. Studies in Fuzziness and Soft Computing, Springer, **207**, 439–445 (2006)
29. Lindsten, F., Ohlsson, H., Ljung L.: Just relax and come clustering! A convexication of k-means clustering. tech. report, Linköpings universitet (2011)
30. Mordukhovich, B.S., Nam, N.M., Salinas, J.: Solving a generalized Heron problem by means of convex analysis. Amer. Math. Monthly, **119**(2), 87–99 (2012)

31. Mordukhovich, B.S., Nam, N.M., Salinas, J.: Applications of variational analysis to a generalized Heron problem. Appl. Anal. **91**(10), 1915–1942 (2012)
32. Nesterov, Y.: Smooth minimization of non-smooth functions. Math. Program. **103**(1), 127–152 (2005)
33. Nesterov, Y.: A method for unconstrained convex minimization problem with the rate of convergence $\mathscr{O}(1/k^2)$. Doklady AN SSSR (translated as Soviet Math. Docl.), **269**, 543–547 (1983)
34. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pacific J. Math **33**(1), 209–216 (1970)
35. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optimization **14**(5), 877–898 (1976)
36. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge (2004)
37. Simons, S.: From Hahn-Banach to Monotonicity. Springer, Berlin (2008)
38. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J. Optim. **29**(1), 119–138 (1991)
39. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control Optimization **38**(2), 431–446 (2000)
40. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**(3), 667–681 (2013)
41. Zălinescu, C.: Convex Analysis in General Vector Spaces. World Scientific, Singapore, (2002)