Inertial Douglas-Rachford splitting for monotone inclusion problems

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

January 5, 2015

Abstract. We propose an inertial Douglas-Rachford splitting algorithm for finding the set of zeros of the sum of two maximally monotone operators in Hilbert spaces and investigate its convergence properties. To this end we formulate first the inertial version of the Krasnosel'skiĭ–Mann algorithm for approximating the set of fixed points of a nonexpansive operator, for which we also provide an exhaustive convergence analysis. By using a product space approach we employ these results to the solving of monotone inclusion problems involving linearly composed and parallel-sum type operators and provide in this way iterative schemes where each of the maximally monotone mappings is accessed separately via its resolvent. We consider also the special instance of solving a primal-dual pair of nonsmooth convex optimization problems and illustrate the theoretical results via some numerical experiments in clustering and location theory.

Key Words. inertial splitting algorithm, Douglas–Rachford splitting, Krasnosel'skiĭ–Mann algorithm, primal-dual algorithm, convex optimization

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

1 Introduction and preliminaries

The problem of approaching the set of zeros of maximally monotone operators by means of splitting iterative algorithms, where each of the operators involved is evaluated separately, either via its resolvent in the set-valued case, or by means of the operator itself in the single-valued case, continues to be a very attractive research area. This is due to its applicability in the context of solving real-life problems which can be modeled as nondifferentiable convex optimization problems, like those arising in image processing, signal recovery, support vector machines classification, location theory, clustering, network communications, etc.

In this manuscript we focus our attention on the *Douglas-Rachford algorithm* which approximates the set of zeros of the sum of two maximally monotone operators. Introduced in [25] in the particular context of solving matrix equations, its convergence properties have been investigated also in [28]. One of the methods for proving the convergence of the classical Douglas–Rachford splitting method is by treating it as a particular case of the Krasnosel'skiĭ–Mann algorithm designed for finding fixed points of nonexpansive operators (see [6], [22]). This approach has the advantage of allowing the inclusion of relaxation parameters in the update rules of the iterates.

In this paper we introduce and investigate the convergence properties of an inertial Douglas-Rachford splitting algorithm. Inertial proximal methods go back to [1,3], where it has been noticed that the discretization of a differential system of second-order in time gives rise to a generalization of the classical proximal-point algorithm for finding the zeros of a maximally monotone operator (see [38]), nowadays called *inertial proximal-point algorithm*. One of the main features of the inertial proximal algorithm is that the next iterate is defined by making use of the last two iterates. Since its introduction, one can notice an increasing interest in algorithms having this particularity, see [2–4, 11, 17, 19, 20, 29, 30, 33, 35, 36].

In order to prove the convergence of the proposed inertial Douglas-Rachford algorithm we formulate first the inertial version of the Krasnosel'skiĭ–Mann algorithm for approximating the set of fixed points of a nonexpansive operator and investigate its convergence properties. The convergence of the inertial Douglas-Rachford scheme is then derived by applying the inertial version of the Krasnosel'skiĭ–Mann algorithm to the composition of the reflected resolvents of the maximally monotone operators involved in the monotone inclusion problem.

The second major aim of the paper is to make use of these results when formulating an *inertial Douglas-Rachford primal-dual algorithm* designed to solve monotone inclusion problems involving linearly composed and parallel-sum type operators. Let us mention that the classical Douglas-Rachford algorithm cannot handle monotone inclusion problems where some of the set-valued mappings involved are composed with linear continuous operators, since in general there is no closed form for the resolvent of the composition. The same applies in the case of monotone inclusion problems involving parallel-sum type operators. *Primal-dual methods* are modern techniques which overcome this drawback. We underline further highlights of these methods, like their full decomposability and the fact that they are able to solve concomitantly a primal monotone inclusion problem and its dual one in the sense of Attouch-Théra [5] (see [12–16, 18, 23, 24, 41] for more details).

The structure of the paper is the following. In the remainder of this section we recall some elements of the theory of maximal monotone operators and some convergence results needed in the paper. The next section contains the inertial type of the Krasnosel'skiĭ–Mann scheme followed by the inertial Douglas-Rachford algorithm with corresponding weak and strong convergence results. In Section 3 we formulate the inertial Douglas-Rachford primal-dual splitting algorithm and study its convergence, while in Section 4 we make use of this iterative scheme when solving primal-dual pairs of convex optimization problems. We close the paper by illustrating the theoretical results via some numerical experiments in clustering and location theory.

For the notions and results presented as follows we refer the reader to [6,8,9,26,39,42]. Let $\mathbb{N} = \{0, 1, 2, ...\}$ be the set of nonnegative integers. Let $\mathcal{H}$ be a real Hilbert space with *inner product* $\langle \cdot, \cdot \rangle$ and associated *norm* $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. The symbols $\rightharpoonup$ and $\rightarrow$ denote

2

weak and strong convergence, respectively. The following identity will be used several times in the paper (see [6, Corollary 2.14]):

$$\|\alpha x + (1-\alpha)y\|^2 + \alpha(1-\alpha)\|x-y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 \ \forall \alpha \in \mathbb{R} \ \forall(x,y) \in \mathcal{H} \times \mathcal{H}. \tag{1}$$

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

For an arbitrary set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ we denote by $\operatorname{Gr} A = \{(x,u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ its *graph*, by $\operatorname{dom} A = \{x \in \mathcal{H} : Ax \neq \varnothing\}$ its *domain*, by $\operatorname{ran} A = \cup_{x \in \mathcal{H}} Ax$ its *range* and by $A^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ its *inverse operator*, defined by $(u,x) \in \operatorname{Gr} A^{-1}$ if and only if $(x,u) \in \operatorname{Gr} A$. We use also the notation $\operatorname{zer} A = \{x \in \mathcal{H} : 0 \in Ax\}$ for the *set of zeros* of A. We say that A is *monotone* if $\langle x-y, u-v \rangle \geq 0$ for all $(x,u),(y,v) \in \operatorname{Gr} A$. A monotone operator A is said to be *maximally monotone*, if there exists no proper monotone extension of the graph of A on $\mathcal{H} \times \mathcal{H}$. The *resolvent* of A is

$$J_A : \mathcal{H} \rightrightarrows \mathcal{H}, J_A = (\operatorname{Id}_{\mathcal{H}} + A)^{-1},$$

and the *reflected resolvent* of A is

$$R_A : \mathcal{H} \rightrightarrows \mathcal{H}, R_A = 2J_A - \operatorname{Id}_{\mathcal{H}},$$

where $\operatorname{Id}_{\mathcal{H}} : \mathcal{H} \to \mathcal{H}, \operatorname{Id}_{\mathcal{H}}(x) = x$ for all $x \in \mathcal{H}$, is the *identity operator* on $\mathcal{H}$. Moreover, if A is maximally monotone, then $J_A : \mathcal{H} \to \mathcal{H}$ is single-valued and maximally monotone (see [6, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [6, Proposition 23.2])

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

and (see [6, Proposition 23.18])

$$J_{\gamma A} + \gamma J_{\gamma^{-1}A^{-1}} \circ \gamma^{-1} \operatorname{Id}_{\mathcal{H}} = \operatorname{Id}_{\mathcal{H}}. \tag{2}$$

Further, let us mention some classes of operators that are used in the paper. The operator A is said to be *uniformly monotone* if there exists an increasing function $\phi_A : [0,+\infty) \to [0,+\infty]$ that vanishes only at 0, and

$$\langle x-y, u-v \rangle \geq \phi_A(\|x-y\|) \ \forall(x,u),(y,v) \in \operatorname{Gr} A. \tag{3}$$

Prominent representatives of the class of uniformly monotone operators are the strongly monotone operators. Let $\gamma > 0$ be arbitrary. We say that A is γ-*strongly monotone*, if $\langle x-y, u-v \rangle \geq \gamma\|x-y\|^2$ for all $(x,u),(y,v) \in \operatorname{Gr} A$.

We consider also the class of nonexpansive operators. An operator $T : D \to \mathcal{H}$, where $D \subseteq \mathcal{H}$ is nonempty, is said to be *nonexpansive*, if $\|Tx - Ty\| \leq \|x-y\|$ for all $x,y \in D$. We use the notation $\operatorname{Fix} T = \{x \in D : Tx = x\}$ for the set of *fixed points* of T. Let us mention that the resolvent and the reflected resolvent of a maximally monotone operator are both nonexpansive (see [6, Corollary 23.10]).

The following result, which is a consequence of the demiclosedness principle (see [6, Theorem 4.17]), will be useful in the proof of the convergence of the inertial version of the Krasnosel'skiĭ–Mann algorithm.

3

Lemma 1 *(see [6, Corollary 4.18]) Let $D \subseteq \mathcal{H}$ be nonempty closed and convex, $T : D \to \mathcal{H}$ be nonexpansive and let $(x_n)_{n \in \mathbb{N}}$ be a sequence in D and $x \in \mathcal{H}$ such that $x_n \rightharpoonup x$ and $Tx_n - x_n \to 0$ as $n \to +\infty$. Then $x \in \operatorname{Fix} T$.*

The *parallel sum* of two operators $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined by $A \,\square\, B : \mathcal{H} \rightrightarrows \mathcal{H}, A \,\square\, B = (A^{-1} + B^{-1})^{-1}$. If A and B are monotone, then we have the following characterization of the set of zeros of their sum (see [6, Proposition 25.1(ii)]):

$$\operatorname{zer}(A + B) = J_{\gamma B}(\operatorname{Fix} R_{\gamma A} R_{\gamma B}) \ \forall \gamma > 0. \tag{4}$$

The following result is a direct consequence of [6, Corollary 25.5] and will be used in the proof of the convergence of the inertial Douglas–Rachford splitting algorithm.

Lemma 2 *Let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone operators and the sequences $(x_n, u_n)_{n \in \mathbb{N}} \in \operatorname{Gr} A$, $(y_n, v_n)_{n \in \mathbb{N}} \in \operatorname{Gr} B$ such that $x_n \rightharpoonup x, u_n \rightharpoonup u, y_n \rightharpoonup y, v_n \rightharpoonup v$, $u_n + v_n \to 0$ and $x_n - y_n \to 0$ as $n \to +\infty$. Then $x = y \in \operatorname{zer}(A + B)$, $(x, u) \in \operatorname{Gr} A$ and $(y, v) \in \operatorname{Gr} B$.*

We close this section by presenting two convergence results which will be crucial for the proof of the main results in the next section.

Lemma 3 *(see [2, Lemma 3]) Let $(\varphi_n)_{n \in \mathbb{N}}, (\delta_n)_{n \in \mathbb{N}}$ and $(\alpha_n)_{n \in \mathbb{N}}$ be sequences in $[0, +\infty)$ such that $\varphi_{n+1} \leq \varphi_n + \alpha_n(\varphi_n - \varphi_{n-1}) + \delta_n$ for all $n \geq 1$, $\sum_{n \in \mathbb{N}} \delta_n < +\infty$ and there exists a real number α with $0 \leq \alpha_n \leq \alpha < 1$ for all $n \in \mathbb{N}$. Then the following hold:*

(i) $\sum_{n \geq 1} [\varphi_n - \varphi_{n-1}]_+ < +\infty$, where $[t]_+ = \max\{t, 0\}$;

(ii) there exists $\varphi^* \in [0, +\infty)$ such that $\lim_{n \to +\infty} \varphi_n = \varphi^*$.

Finally, we recall a well known result on weak convergence in Hilbert spaces.

Lemma 4 *(Opial) Let C be a nonempty set of $\mathcal{H}$ and $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$ such that the following two conditions hold:*

(a) for every $x \in C$, $\lim_{n \to +\infty} \|x_n - x\|$ exists;

(b) every sequential weak cluster point of $(x_n)_{n \in \mathbb{N}}$ is in C.

Then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in C.

2 An inertial Douglas–Rachford splitting algorithm

This section is dedicated to the formulation of an inertial Douglas–Rachford splitting algorithm for finding the set of zeros of the sum of two maximally monotone operators and to the investigation of its convergence properties.

In the first part we propose an inertial version of the Krasnosel'skiĭ–Mann algorithm for approximating the set of fixed points of a nonexpansive operator, a result which has its own interest. Notice that due to the presence of affine combinations in the iterative scheme, we have to restrict the setting to nonexpansive operators defined on affine subspaces. Let us underline that this assumption is fulfilled when considering the composition of

the reflected resolvents of maximally monotone operators, which will be the case for the inertial Douglas–Rachford algorithm. Let us also mention that some inertial versions of the Krasnosel'skiĭ–Mann algorithm have been proposed also in [29], which, however, in order to ensure the convergence of the generated sequence of iterates, ask for a summability condition formulated in terms of this sequence.

Theorem 5 *Let M be a nonempty closed affine subset of $\mathcal{H}$ and $T : M \to M$ a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$. We consider the following iterative scheme:*

$$x_{n+1} = x_n + \alpha_n(x_n - x_{n-1}) + \lambda_n \Big[T\big(x_n + \alpha_n(x_n - x_{n-1})\big) - x_n - \alpha_n(x_n - x_{n-1}) \Big] \ \forall n \geq 1 \ (5)$$

where x_0, x_1 are arbitrarily chosen in M, $(\alpha_n)_{n\geq 1}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ are such that

$$\delta > \frac{\alpha^2(1+\alpha) + \alpha\sigma}{1 - \alpha^2} \ \text{and} \ 0 < \lambda \leq \lambda_n \leq \frac{\delta - \alpha\Big[\alpha(1+\alpha) + \alpha\delta + \sigma\Big]}{\delta\Big[1 + \alpha(1+\alpha) + \alpha\delta + \sigma\Big]} \ \forall n \geq 1. \quad (6)$$

Then the following statements are true:

(i) $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$;

(ii) $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

Proof. Let us start with the remark that, due to the choice of δ, $\lambda_n \in (0,1)$ for every $n \geq 1$. Furthermore, we would like to notice that, since M is affine, the iterative scheme provides a well-defined sequence in M.

(i) We denote

$$w_n := x_n + \alpha_n(x_n - x_{n-1}) \ \forall n \geq 1.$$

Then the iterative scheme reads for every $n \geq 1$:

$$x_{n+1} = w_n + \lambda_n(Tw_n - w_n). \quad (7)$$

Let us fix an element $y \in \operatorname{Fix} T$ and $n \geq 1$. It follows from (1) and the nonexpansiveness of T that

$$\|x_{n+1} - y\|^2 = (1 - \lambda_n)\|w_n - y\|^2 + \lambda_n\|Tw_n - Ty\|^2 - \lambda_n(1 - \lambda_n)\|Tw_n - w_n\|^2$$
$$\leq \|w_n - y\|^2 - \lambda_n(1 - \lambda_n)\|Tw_n - w_n\|^2. \quad (8)$$

Applying (1) again, we have

$$\|w_n - y\|^2 = \|(1 + \alpha_n)(x_n - y) - \alpha_n(x_{n-1} - y)\|^2$$
$$= (1 + \alpha_n)\|x_n - y\|^2 - \alpha_n\|x_{n-1} - y\|^2 + \alpha_n(1 + \alpha_n)\|x_n - x_{n-1}\|^2,$$

hence from (8) we obtain

$$\|x_{n+1} - y\|^2 - (1 + \alpha_n)\|x_n - y\|^2 + \alpha_n\|x_{n-1} - y\|^2 \leq -\lambda_n(1 - \lambda_n)\|Tw_n - w_n\|^2$$
$$+ \alpha_n(1 + \alpha_n)\|x_n - x_{n-1}\|^2. \quad (9)$$

Furthermore, we have

$$\|Tw_n - w_n\|^2 = \left\|\frac{1}{\lambda_n}(x_{n+1} - x_n) + \frac{\alpha_n}{\lambda_n}(x_{n-1} - x_n)\right\|^2$$

$$= \frac{1}{\lambda_n^2}\|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2}{\lambda_n^2}\|x_n - x_{n-1}\|^2 + 2\frac{\alpha_n}{\lambda_n^2}\langle x_{n+1} - x_n, x_{n-1} - x_n\rangle$$

$$\geq \frac{1}{\lambda_n^2}\|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2}{\lambda_n^2}\|x_n - x_{n-1}\|^2$$

$$+ \frac{\alpha_n}{\lambda_n^2}\left(-\rho_n\|x_{n+1} - x_n\|^2 - \frac{1}{\rho_n}\|x_n - x_{n-1}\|^2\right), \tag{10}$$

where we denote $\rho_n := \frac{1}{\alpha_n + \delta\lambda_n}$.

We derive from (9) and (10) the inequality

$$\|x_{n+1} - y\|^2 - (1 + \alpha_n)\|x_n - y\|^2 + \alpha_n\|x_{n-1} - y\|^2 \leq \frac{(1 - \lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n}\|x_{n+1} - x_n\|^2$$

$$+ \gamma_n\|x_n - x_{n-1}\|^2, \tag{11}$$

where

$$\gamma_n := \alpha_n(1 + \alpha_n) + \alpha_n(1 - \lambda_n)\frac{1 - \rho_n\alpha_n}{\rho_n\lambda_n} > 0, \tag{12}$$

since $\rho_n\alpha_n < 1$ and $\lambda_n \in (0, 1)$.

Again, taking into account the choice of ρ_n we have

$$\delta = \frac{1 - \rho_n\alpha_n}{\rho_n\lambda_n}$$

and from (12), it follows

$$\gamma_n = \alpha_n(1 + \alpha_n) + \alpha_n(1 - \lambda_n)\delta \leq \alpha(1 + \alpha) + \alpha\delta \ \forall n \geq 1. \tag{13}$$

In the following we use some techniques from [3] adapted to our setting. We define the sequences $\varphi_n := \|x_n - y\|^2$ for all $n \in \mathbb{N}$ and $\mu_n := \varphi_n - \alpha_n\varphi_{n-1} + \gamma_n\|x_n - x_{n-1}\|^2$ for all $n \geq 1$. Using the monotonicity of $(\alpha_n)_{n\geq 1}$ and the fact that $\varphi_n \geq 0$ for all $n \in \mathbb{N}$, we get

$$\mu_{n+1} - \mu_n \leq \varphi_{n+1} - (1 + \alpha_n)\varphi_n + \alpha_n\varphi_{n-1} + \gamma_{n+1}\|x_{n+1} - x_n\|^2 - \gamma_n\|x_n - x_{n-1}\|^2.$$

Employing (11), we have

$$\mu_{n+1} - \mu_n \leq \left(\frac{(1 - \lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1}\right)\|x_{n+1} - x_n\|^2 \ \forall n \geq 1. \tag{14}$$

We claim that

$$\frac{(1 - \lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1} \leq -\sigma \ \forall n \geq 1. \tag{15}$$

Let be $n \geq 1$. Indeed, by the choice of ρ_n, we get

$$\frac{(1 - \lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1} \leq -\sigma$$

$$\Longleftrightarrow \lambda_n(\gamma_{n+1} + \sigma) + (\alpha_n\rho_n - 1)(1 - \lambda_n) \leq 0$$

$$\Longleftrightarrow \lambda_n(\gamma_{n+1} + \sigma) - \frac{\delta\lambda_n(1 - \lambda_n)}{\alpha_n + \delta\lambda_n} \leq 0$$

$$\Longleftrightarrow (\alpha_n + \delta\lambda_n)(\gamma_{n+1} + \sigma) + \delta\lambda_n \leq \delta.$$

6

By using (13), we have

$$(\alpha_n + \delta\lambda_n)(\gamma_{n+1} + \sigma) + \delta\lambda_n \leq (\alpha + \delta\lambda_n)\Big(\alpha(1 + \alpha) + \alpha\delta + \sigma\Big) + \delta\lambda_n \leq \delta,$$

where the last inequality follows by using the upper bound for $(\lambda_n)_{n\geq 1}$ in (6). Hence the claim in (15) is true.

We obtain from (14) and (15) that

$$\mu_{n+1} - \mu_n \leq -\sigma\|x_{n+1} - x_n\|^2 \;\; \forall n \geq 1. \tag{16}$$

The sequence $(\mu_n)_{n\geq 1}$ is nonincreasing and the bound for $(\alpha_n)_{n\geq 1}$ delivers

$$-\alpha\varphi_{n-1} \leq \varphi_n - \alpha\varphi_{n-1} \leq \mu_n \leq \mu_1 \;\; \forall n \geq 1. \tag{17}$$

We obtain

$$\varphi_n \leq \alpha^n\varphi_0 + \mu_1 \sum_{k=0}^{n-1} \alpha^k \leq \alpha^n\varphi_0 + \frac{\mu_1}{1 - \alpha} \;\; \forall n \geq 1,$$

where we notice that $\mu_1 = \varphi_1 \geq 0$ (due to the relation $\alpha_1 = 0$). Combining (16) and (17), we get for all $n \geq 1$

$$\sigma\sum_{k=1}^{n} \|x_{k+1} - x_k\|^2 \leq \mu_1 - \mu_{n+1} \leq \mu_1 + \alpha\varphi_n \leq \alpha^{n+1}\varphi_0 + \frac{\mu_1}{1 - \alpha},$$

which shows that $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$.

(ii) We prove this by using the result of Opial given in Lemma 4. We have proven above that for an arbitrary $y \in \operatorname{Fix} T$ the inequality (11) is true. By part (i), (13) and Lemma 3 we derive that $\lim_{n\to+\infty} \|x_n - y\|$ exists (we take into consideration also that in (11) $\alpha_n\rho_n < 1$ for all $n \geq 1$). On the other hand, let x be a sequential weak cluster point of $(x_n)_{n\in\mathbb{N}}$, that is, the latter has a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ fulfilling $x_{n_k} \rightharpoonup x$ as $k \to +\infty$. By part (i), the definition of w_n and the upper bound for $(\alpha_n)_{n\geq 1}$, we get $w_{n_k} \rightharpoonup x$ as $k \to +\infty$. Furthermore, from (7) we have

$$\|Tw_n - w_n\| = \frac{1}{\lambda_n}\|x_{n+1} - w_n\| \leq \frac{1}{\lambda}\|x_{n+1} - w_n\| \leq \frac{1}{\lambda}\big(\|x_{n+1} - x_n\| + \alpha\|x_n - x_{n-1}\|\big), \tag{18}$$

thus by (i) we obtain $Tw_{n_k} - w_{n_k} \to 0$ as $k \to +\infty$. Applying now Lemma 1 for the sequence $(w_{n_k})_{k\in\mathbb{N}}$ we conclude that $x \in \operatorname{Fix} T$. Since the two assumptions of Lemma 4 are verified, it follows that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$. ∎

Remark 6 The condition $\alpha_1 = 0$ was imposed in order to ensure $\mu_1 \geq 0$, which is needed in the proof. An alternative is to require that $x_0 = x_1$, in which case the assumption $\alpha_1 = 0$ is not anymore necessary.

Remark 7 Assuming that $\alpha = 0$ (which forces $\alpha_n = 0$ for all $n \geq 1$), the iterative scheme in the previous theorem is nothing else than the one in the classical Krasnosel'skiĭ–Mann algorithm:

$$x_{n+1} = x_n + \lambda_n(Tx_n - x_n) \;\; \forall n \geq 1. \tag{19}$$

Let us mention that the convergence of this iterative scheme can be proved under more general hypotheses, namely when M is a nonempty closed and convex set and the sequence $(\lambda_n)_{n\in\mathbb{N}}$ satisfies the relation $\sum_{n\in\mathbb{N}} \lambda_n(1 - \lambda_n) = +\infty$ (see [6, Theorem 5.14]).

We are now in position to state the inertial Douglas–Rachford splitting algorithm and to present its convergence properties.

Theorem 8 *(Inertial Douglas–Rachford splitting algorithm) Let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone operators such that $\mathrm{zer}(A + B) \neq \varnothing$. Consider the following iterative scheme:*

$$(\forall n \geq 1) \quad \begin{cases} y_n = J_{\gamma B}[x_n + \alpha_n(x_n - x_{n-1})] \\ z_n = J_{\gamma A}[2y_n - x_n - \alpha_n(x_n - x_{n-1})] \\ x_{n+1} = x_n + \alpha_n(x_n - x_{n-1}) + \lambda_n(z_n - y_n) \end{cases}$$

where $\gamma > 0$, x_0, x_1 are arbitrarily chosen in $\mathcal{H}$, $(\alpha_n)_{n \geq 1}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ are such that

$$\delta > \frac{\alpha^2(1 + \alpha) + \alpha\sigma}{1 - \alpha^2} \quad and \quad 0 < \lambda \leq \lambda_n \leq 2\frac{\delta - \alpha\big[\alpha(1 + \alpha) + \alpha\delta + \sigma\big]}{\delta\big[1 + \alpha(1 + \alpha) + \alpha\delta + \sigma\big]} \quad \forall n \geq 1.$$

Then there exists $x \in \mathrm{Fix}(R_{\gamma A} R_{\gamma B})$ such that the following statements are true:

(i) *$J_{\gamma B} x \in \mathrm{zer}(A + B)$;*

(ii) *$\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$;*

(iii) *$(x_n)_{n \in \mathbb{N}}$ converges weakly to x;*

(iv) *$y_n - z_n \to 0$ as $n \to +\infty$;*

(v) *$(y_n)_{n \geq 1}$ converges weakly to $J_{\gamma B} x$;*

(vi) *$(z_n)_{n \geq 1}$ converges weakly to $J_{\gamma B} x$;*

(vii) *if A or B is uniformly monotone, then $(y_n)_{n \geq 1}$ and $(z_n)_{n \geq 1}$ converge strongly to the unique point in $\mathrm{zer}(A + B)$.*

Proof. We use again the notation $w_n = x_n + \alpha_n(x_n - x_{n-1})$ for all $n \geq 1$. Taking into account the iteration rules and the definition of the reflected resolvent, the iterative scheme in the enunciation of the theorem can be written as

$$(\forall n \geq 1) \ x_{n+1} = w_n + \lambda_n\Big[J_{\gamma A} \circ (2J_{\gamma B} - \mathrm{Id})w_n - J_{\gamma B}w_n\Big]$$

$$= w_n + \lambda_n\left[\left(\frac{\mathrm{Id} + R_{\gamma A}}{2} \circ R_{\gamma B}\right)w_n - \frac{\mathrm{Id} + R_{\gamma B}}{2}w_n\right]$$

$$= w_n + \frac{\lambda_n}{2}(Tw_n - w_n), \tag{20}$$

where $T := R_{\gamma A} \circ R_{\gamma B} : \mathcal{H} \to \mathcal{H}$ is a nonexpansive operator. From (4) we have $\mathrm{zer}(A+B) = J_{\gamma B}(\mathrm{Fix}\, T)$, hence $\mathrm{Fix}\, T \neq \varnothing$. By applying Theorem 5, there exists $x \in \mathrm{Fix}\, T$ such that (i)-(iii) hold.

(iv) follows from Theorem 5, (18) and that $z_n - y_n = \frac{1}{2}(Tw_n - w_n)$ for $n \geq 1$.

(v) We will show that $(y_n)_{n \geq 1}$ is bounded and that $J_{\gamma B} x$ is the unique weak sequential cluster point of $(y_n)_{n \geq 1}$. From here the conclusion will automatically follow. By using that $J_{\gamma B}$ is nonexpansive, for all $n \geq 1$ we have

$$\|y_n - y_1\| = \|J_{\gamma B}w_n - J_{\gamma B}w_1\| \leq \|w_n - w_1\| = \|x_n - x_1 + \alpha_n(x_n - x_{n-1})\|.$$

8

Since $(x_n)_{n\in\mathbb{N}}$ is bounded (by (iii)) and $(\alpha_n)_{n\geq 1}$ is also bounded, so is the sequence $(y_n)_{n\geq 1}$.

Now let y be a sequential weak cluster point of $(y_n)_{n\geq 1}$, that is, the latter has a subsequence $(y_{n_k})_{k\in\mathbb{N}}$ fulfilling $y_{n_k} \rightharpoonup y$ as $k \to +\infty$. We use the notations $u_n := 2y_n - w_n - z_n$ and $v_n := w_n - y_n$ for all $n \geq 1$. The definitions of the resolvent yields

$$(z_n, u_n) \in \text{Gr}(\gamma A), \ (y_n, v_n) \in \text{Gr}(\gamma B) \text{ and } u_n + v_n = y_n - z_n \ \forall n \geq 1. \qquad (21)$$

Furthermore, by (ii), (iii) and (iv) we derive

$$z_{n_k} \rightharpoonup y, w_{n_k} \rightharpoonup x, u_{n_k} \rightharpoonup y - x \text{ and } v_{n_k} \rightharpoonup x - y \text{ as } k \to +\infty.$$

Using again (ii) and Lemma 2 we obtain $y \in \text{zer}(\gamma A + \gamma B) = \text{zer}(A+B), (y, y-x) \in \text{Gr}\,\gamma A$ and $(y, x-y) \in \text{Gr}\,\gamma B$. As a consequence, $y = J_{\gamma B}x$.

(vi) Follows from (iv) and (v).

(vii) We prove the statement in case A is uniformly monotone, the situation when B fulfills this condition being similar. Denote $y = J_{\gamma B}x$. There exists an increasing function $\phi_A : [0, +\infty) \to [0, +\infty]$ that vanishes only at 0 such that (see also (21) and the considerations made in the proof of (v))

$$\gamma\phi_A(\|z_n - y\|) \leq \langle z_n - y, u_n - y + x \rangle \ \forall n \geq 1.$$

Moreover, since B is monotone we have (see (21))

$$0 \leq \langle y_n - y, v_n - x + y \rangle = \langle y_n - y, y_n - z_n - u_n - x + y \rangle \ \forall n \geq 1.$$

Summing up the last two inequalities we obtain

$$\gamma\phi_A(\|z_n - y\|) \leq \langle z_n - y_n, u_n - y_n + x \rangle = \langle z_n - y_n, y_n - z_n - w_n + x \rangle \ \forall n \geq 1.$$

Since $z_n - y_n \to 0$ and $w_n \rightharpoonup x$ as $n \to +\infty$, from the last inequality we get $\lim_{n\to+\infty} \phi_A(\|z_n - y\|) = 0$, hence $z_n \to y$ and therefore $y_n \to y$ as $n \to +\infty$. ∎

Remark 9 In case $\alpha = 0$, which forces $\alpha_n = 0$ for all $n \geq 1$, the iterative scheme in Theorem 8 becomes the classical Douglas–Rachford splitting algorithm (see [6, Theorem 25.6]):

$$(\forall n \geq 1) \quad \begin{cases} y_n = J_{\gamma B}x_n \\ z_n = J_{\gamma A}(2y_n - x_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n), \end{cases}$$

the convergence of which holds under the assumption $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$. Let us mention that the weak convergence of the sequence $(y_n)_{n\geq 1}$ to a point in $\text{zer}(A + B)$ has been for the first time reported in [40].

Remark 10 In case $Bx = 0$ for all $x \in \mathcal{H}$, the iterative scheme in Theorem 8 becomes

$$x_{n+1} = \lambda_n J_{\gamma A}\big(x_n + \alpha_n(x_n - x_{n-1})\big) + (1 - \lambda_n)(x_n + \alpha_n(x_n - x_{n-1})) \ \forall n \geq 1,$$

which was already considered in [2] as a proximal-point algorithm (see [38]) in the context of solving the monotone inclusion problem $0 \in Ax$. Notice that in this scheme in each iteration a constant step-size $\gamma > 0$ is considered. Proximal-point algorithms of inertial-type with variable step-sizes have been proposed and investigated, for instance, in [3, Theorem 2.1], [2] and [11, Remark 7].

3 Solving monotone inclusion problems involving mixtures of linearly composed and parallel-sum type operators

In this section we apply the inertial Douglas–Rachford algorithm proposed in Section 2 to a highly structured primal-dual system of monotone inclusions by making use of appropriate splitting techniques. The problem under investiagtion reads as follows.

Problem 11 Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and let $z \in \mathcal{H}$. Moreover, let m be a strictly positive integer and for every $i \in \{1,...,m\}$, let $r_i \in \mathcal{G}_i$, $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ and $D_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be maximally monotone operators and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be nonzero linear continuous operators. The problem is to solve the primal inclusion

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^*(B_i \,\square\, D_i)(L_i\overline{x} - r_i) \tag{22}$$

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in Ax \\ \overline{v}_i \in (B_i \,\square\, D_i)(L_ix - r_i), \ i = 1,...,m. \end{cases} \tag{23}$$

We say that $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 ... \times \mathcal{G}_m$ is a primal-dual solution to Problem 11, if

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in A\overline{x} \text{ and } \overline{v}_i \in (B_i \,\square\, D_i)(L_i\overline{x} - r_i), \ i = 1,...,m. \tag{24}$$

Note that, if $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 ... \times \mathcal{G}_m$ is a primal-dual solution to Problem 11, then $\overline{x}$ is a solution to (22) and $(\overline{v}_1,..., \overline{v}_m)$ is a solution to (23). On the other hand, if $\overline{x} \in \mathcal{H}$ is a solution to (22), then there exists $(\overline{v}_1,..., \overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$ such that $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ is a primal-dual solution to Problem 11. Equivalently, if $(\overline{v}_1,..., \overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a solution to (23), then there exists $\overline{x} \in \mathcal{H}$ such that $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ is a primal-dual solution to Problem 11.

Several particular instances of the primal-dual system of monotone inclusions (22)–(23) when applied to convex optimization problems can be found in [23, 41].

The inertial primal-dual Douglas-Rachford algorithm we would like to propose for solving (22)–(23) is formulated as follows.

Algorithm 12 Let $x_0, x_1 \in \mathcal{H}$, $v_{i,0}, v_{i,1} \in \mathcal{G}_i$, $i = 1,...,m$, and $\tau, \sigma_i > 0$, $i = 1,...,m$, be such that

$$\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < 4.$$

Furthermore, let $(\alpha_n)_{n \geq 1}$ be a nondecreasing sequence with $\alpha_1 = 0$ and $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ and the sequence $(\lambda_n)_{n \geq 1}$ be such that

$$\delta > \frac{\alpha^2(1+\alpha) + \alpha\sigma}{1 - \alpha^2} \text{ and } 0 < \lambda \leq \lambda_n \leq 2\frac{\delta - \alpha\left[\alpha(1+\alpha) + \alpha\delta + \sigma\right]}{\delta\left[1 + \alpha(1+\alpha) + \alpha\delta + \sigma\right]} \ \forall n \geq 1.$$

Set

$$(\forall n \geq 1) \quad \begin{vmatrix} p_{1,n} = J_{\tau A}\left(x_n + \alpha_n(x_n - x_{n-1}) - \frac{\tau}{2}\sum_{i=1}^m L_i^*(v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1})) + \tau z\right) \\ w_{1,n} = 2p_{1,n} - x_n - \alpha_n(x_n - x_{n-1}) \\ \text{For } i = 1,...,m \\ \quad \begin{vmatrix} p_{2,i,n} = J_{\sigma_i B_i^{-1}}\left(v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1}) + \frac{\sigma_i}{2}L_i w_{1,n} - \sigma_i r_i\right) \\ w_{2,i,n} = 2p_{2,i,n} - v_{i,n} - \alpha_n(v_{i,n} - v_{i,n-1}) \end{vmatrix} \\ z_{1,n} = w_{1,n} - \frac{\tau}{2}\sum_{i=1}^m L_i^* w_{2,i,n} \\ x_{n+1} = x_n + \alpha_n(x_n - x_{n-1}) + \lambda_n(z_{1,n} - p_{1,n}) \\ \text{For } i = 1,...,m \\ \quad \begin{vmatrix} z_{2,i,n} = J_{\sigma_i D_i^{-1}}\left(w_{2,i,n} + \frac{\sigma_i}{2}L_i(2z_{1,n} - w_{1,n})\right) \\ v_{i,n+1} = v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1}) + \lambda_n(z_{2,i,n} - p_{2,i,n}). \end{vmatrix} \end{vmatrix} \tag{25}$$

Theorem 13 *In Problem 11, suppose that*

$$z \in \mathrm{ran}\left(A + \sum_{i=1}^m L_i^*(B_i \square D_i)(L_i \cdot - r_i)\right), \tag{26}$$

and consider the sequences generated by Algorithm 12. Then there exists $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 ... \times \mathcal{G}_m$ such that the following statements are true:

(i) *By setting*

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

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

the element $(\overline{p}_1, \overline{p}_{2,1},..., \overline{p}_{2,m}) \in \mathcal{H} \times \mathcal{G}_1 \times... \times \mathcal{G}_m$ is a primal-dual solution to Problem 11;

(ii) $\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$ *and* $\sum_{n \in \mathbb{N}} \|v_{i,n+1} - v_{i,n}\|^2 < +\infty$, $i = 1,...,m$;

(iii) $(x_n, v_{1,n},..., v_{m,n})_{n \in \mathbb{N}}$ *converges weakly to* $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$;

(iv) $(p_{1,n} - z_{1,n}, p_{2,1,n} - z_{2,1,n},..., p_{2,m,n} - z_{2,m,n}) \to 0$ *as* $n \to +\infty$;

(v) $(p_{1,n}, p_{2,1,n},..., p_{2,m,n})_{n \geq 1}$ *converges weakly to* $(\overline{p}_1, \overline{p}_{2,1},..., \overline{p}_{2,m})$;

(vi) $(z_{1,n}, z_{2,1,n},..., z_{2,m,n})_{n \geq 1}$ *converges weakly to* $(\overline{p}_1, \overline{p}_{2,1},..., \overline{p}_{2,m})$;

(vii) *if A and B_i^{-1}, $i = 1,...,m$, are uniformly monotone, then $(p_{1,n}, p_{2,1,n},..., p_{2,m,n})_{n \geq 1}$ and $(z_{1,n}, z_{2,1,n},..., z_{2,m,n})_{n \geq 1}$ converge strongly to the unique primal-dual solution $(\overline{p}_1, \overline{p}_{2,1},..., \overline{p}_{2,m})$ to Problem 11.*

Proof. For the proof we use Theorem 8 and adapt the techniques from [15] (see also [41]) to the given settings. We consider the Hilbert space $\mathcal{K} = \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ endowed with

inner product and associated norm defined, for $(x, v_1,..., v_m), (y, q_1,..., q_m) \in \mathcal{K}$, via

$$\langle (x, v_1,..., v_m), (y, q_1,..., q_m) \rangle_{\mathcal{K}} = \langle x, y \rangle_{\mathcal{H}} + \sum_{i=1}^{m} \langle v_i, q_i \rangle_{\mathcal{G}_i}$$

$$\text{and } \|(x, v_1,..., v_m)\|_{\mathcal{K}} = \sqrt{\|x\|_{\mathcal{H}}^2 + \sum_{i=1}^{m} \|v_i\|_{\mathcal{G}_i}^2}, \tag{27}$$

respectively. Furthermore, we consider the set-valued operator

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

which is maximally monotone, since A and B_i, $i = 1,..., m$, are maximally monotone (see [6, Proposition 20.22 and Proposition 20.23]), and the linear continuous operator

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

which is skew-symmetric (i. e. $\boldsymbol{S}^* = -\boldsymbol{S}$) and hence maximally monotone (see [6, Example 20.30]). Moreover, we consider the set-valued operator

$$\boldsymbol{Q} : \mathcal{K} \rightrightarrows \mathcal{K}, \quad (x, v_1,..., v_m) \mapsto \left(0, D_1^{-1} v_1,..., D_m^{-1} v_m \right),$$

which is once again maximally monotone, since D_i is maximally monotone for $i = 1,..., m$. Therefore, since dom $\boldsymbol{S} = \mathcal{K}$, both $\frac{1}{2}\boldsymbol{S} + \boldsymbol{Q}$ and $\frac{1}{2}\boldsymbol{S} + \boldsymbol{M}$ are maximally monotone (see [6, Corollary 24.4(i)]). Furthermore, one can easily notice that

$$(26) \Leftrightarrow \operatorname{zer}(\boldsymbol{M} + \boldsymbol{S} + \boldsymbol{Q}) \neq \varnothing$$

and

$$(x, v_1,..., v_m) \in \operatorname{zer}(\boldsymbol{M} + \boldsymbol{S} + \boldsymbol{Q})$$
$$\Rightarrow (x, v_1,..., v_m) \text{ is a primal-dual solution to Problem 11.} \tag{28}$$

We also introduce the linear continuous operator

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

which is self-adjoint and ρ-strongly positive (see [15]) for

$$\rho := \left(1 - \frac{1}{2} \sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2} \right) \min \left\{ \frac{1}{\tau}, \frac{1}{\sigma_1},..., \frac{1}{\sigma_m} \right\} > 0,$$

namely, the following inequality holds

$$\langle \boldsymbol{x}, \boldsymbol{V} \boldsymbol{x} \rangle_{\mathcal{K}} \geq \rho \|\boldsymbol{x}\|_{\mathcal{K}}^2 \ \forall \boldsymbol{x} \in \mathcal{K}.$$

Therefore, its inverse operator $\boldsymbol{V}^{-1}$ exists and it fulfills $\|\boldsymbol{V}^{-1}\| \leq \frac{1}{\rho}$.

12

Note that the algorithmic scheme (25) is equivalent to

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

(29)

By introducing the sequences

$$\boldsymbol{x}_n = (x_n, v_{1,n},..., v_{m,n}), \ \boldsymbol{y}_n = (p_{1,n}, p_{2,1,n},..., p_{2,m,n}), \ \boldsymbol{z}_n = (z_{1,n}, z_{2,1,n},..., z_{2,m,n}) \ \forall n \geq 1,$$

the scheme (29) can equivalently be written in the form

$$(\forall n \geq 1) \quad \left|
\begin{array}{l}
\boldsymbol{V}(\boldsymbol{x}_n - \boldsymbol{y}_n + \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})) \in \left(\frac{1}{2}\boldsymbol{S} + \boldsymbol{M}\right)\boldsymbol{y}_n \\
\boldsymbol{V}(2\boldsymbol{y}_n - \boldsymbol{x}_n - \boldsymbol{z}_n - \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})) \in \left(\frac{1}{2}\boldsymbol{S} + \boldsymbol{Q}\right)\boldsymbol{z}_n \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1}) + \lambda_n(\boldsymbol{z}_n - \boldsymbol{y}_n),
\end{array}\right.$$

(30)

which is equivalent to

$$(\forall n \geq 1) \quad \left|
\begin{array}{l}
\boldsymbol{y}_n = \left(\mathrm{Id} + \boldsymbol{V}^{-1}(\frac{1}{2}\boldsymbol{S} + \boldsymbol{M})\right)^{-1}(\boldsymbol{x}_n + \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})) \\
\boldsymbol{z}_n = \left(\mathrm{Id} + \boldsymbol{V}^{-1}(\frac{1}{2}\boldsymbol{S} + \boldsymbol{Q})\right)^{-1}(2\boldsymbol{y}_n - \boldsymbol{x}_n - \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})) \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1}) + \lambda_n(\boldsymbol{z}_n - \boldsymbol{y}_n),
\end{array}\right.$$

(31)

In the following, we consider the Hilbert space $\mathcal{K}_{\boldsymbol{V}}$ with inner product and norm respectively defined, for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{K}$, via

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{K}_{\boldsymbol{V}}} = \langle \boldsymbol{x}, \boldsymbol{V}\boldsymbol{y} \rangle_{\mathcal{K}} \ \text{and} \ \|\boldsymbol{x}\|_{\mathcal{K}_{\boldsymbol{V}}} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{V}\boldsymbol{x} \rangle_{\mathcal{K}}}.$$

(32)

As the set-valued operators $\frac{1}{2}\boldsymbol{S} + \boldsymbol{M}$ and $\frac{1}{2}\boldsymbol{S} + \boldsymbol{Q}$ are maximally monotone on $\mathcal{K}$, the operators

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

(33)

are maximally monotone on $\mathcal{K}_{\boldsymbol{V}}$. Moreover, since $\boldsymbol{V}$ is self-adjoint and ρ-strongly positive, weak and strong convergence in $\mathcal{K}_{\boldsymbol{V}}$ are equivalent with weak and strong convergence in $\mathcal{K}$, respectively.

Taking this into account, it shows that (31) becomes

$$(\forall n \geq 1) \quad \left|
\begin{array}{l}
\boldsymbol{y}_n = J_{\boldsymbol{B}}(\boldsymbol{x}_n + \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})) \\
\boldsymbol{z}_n = J_{\boldsymbol{A}}(2\boldsymbol{y}_n - \boldsymbol{x}_n - \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})) \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1}) + \lambda_n(\boldsymbol{z}_n - \boldsymbol{y}_n),
\end{array}\right.$$

(34)

13

which is the inertial Douglas–Rachford algorithm presented in Theorem 8 in the space $\mathcal{K}_{\boldsymbol{V}}$ for $\gamma = 1$. Furthermore, we have

$$\mathrm{zer}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{zer}(\boldsymbol{V}^{-1}(\boldsymbol{M} + \boldsymbol{S} + \boldsymbol{Q})) = \mathrm{zer}(\boldsymbol{M} + \boldsymbol{S} + \boldsymbol{Q}).$$

(i) By Theorem 8 (i), there exists $\overline{\boldsymbol{x}} = (\overline{x}, \overline{v}_1, ..., \overline{v}_m) \in \mathrm{Fix}(R_{\boldsymbol{A}}R_{\boldsymbol{B}})$, such that $J_{\boldsymbol{B}}\overline{\boldsymbol{x}} \in \mathrm{zer}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{zer}(\boldsymbol{M} + \boldsymbol{S} + \boldsymbol{Q})$. The claim follows from (28) and by identifying $J_{\boldsymbol{B}}\overline{\boldsymbol{x}}$.

(ii) Since $\boldsymbol{V}$ is ρ-strongly positive, we obtain from Theorem 8 (ii) that

$$\rho \sum_{n \in \mathbb{N}} \|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\|_{\mathcal{K}}^2 \leq \sum_{n \in \mathbb{N}} \|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\|_{\mathcal{K}_{\boldsymbol{V}}}^2 < +\infty,$$

and therefore the claim follows by considering (27).

(iii)–(vi) Follows directly from Theorem 8 (iii)–(vi).

(vii) The uniform monotonicity of A and B_i^{-1}, $i = 1,...,m$, implies uniform monotonicity of $\boldsymbol{M}$ on $\mathcal{K}$ (see, for instance, [15, Theorem 2.1 (ii)]), while this further implies uniform monotonicity of $\boldsymbol{B}$ on $\mathcal{K}_{\boldsymbol{V}}$. Therefore, the claim follows from Theorem 8 (vii). ■

4 Convex optimization problems

The aim of this section is to show how the inertial Douglas-Rachford primal-dual algorithm can be implemented when solving a primal-dual pair of convex optimization problems.

We recall first some notations used in the variational case, see [6, 8, 9, 26, 39, 42]. For a function $f : \mathcal{H} \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ is the extended real line, we denote by $\mathrm{dom}\, f = \{x \in \mathcal{H} : f(x) < +\infty\}$ its *effective domain* and say that f is *proper* if $\mathrm{dom}\, f \neq \varnothing$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. We denote by $\Gamma(\mathcal{H})$ the family of proper, convex and lower semi-continuous extended real-valued functions defined on $\mathcal{H}$. Let $f^* : \mathcal{H} \to \overline{\mathbb{R}}$, $f^*(u) = \sup_{x \in \mathcal{H}}\{\langle u, x \rangle - f(x)\}$ for all $u \in \mathcal{H}$, be the *conjugate function* of f. The *subdifferential* of f at $x \in \mathcal{H}$, with $f(x) \in \mathbb{R}$, is the set $\partial f(x) := \{v \in \mathcal{H} : f(y) \geq f(x) + \langle v, y - x \rangle \, \forall y \in \mathcal{H}\}$. We take by convention $\partial f(x) := \varnothing$, if $f(x) \in \{\pm\infty\}$. Notice that if $f \in \Gamma(\mathcal{H})$, then ∂f is a maximally monotone operator (see [37]) and it holds $(\partial f)^{-1} = \partial f^*$. For two proper functions $f, g : \mathcal{H} \to \overline{\mathbb{R}}$, we consider their *infimal convolution*, which is the function $f \Box g : \mathcal{H} \to \overline{\mathbb{R}}$, defined by $(f \Box g)(x) = \inf_{y \in \mathcal{H}}\{f(y) + g(x - y)\}$, for all $x \in \mathcal{H}$.

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

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

$$\inf_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}. \tag{35}$$

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

$$\mathrm{prox}_{\gamma f} + \gamma \, \mathrm{prox}_{(1/\gamma)f^*} \circ \gamma^{-1} \, \mathrm{Id}_{\mathcal{H}} = \mathrm{Id}_{\mathcal{H}}. \tag{36}$$

14

Let us also recall that a proper function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is said to be *uniformly convex*, if there exists an increasing function $\phi : [0, +\infty) \to [0, +\infty]$ which vanishes only at 0 and such that

$$f(tx + (1-t)y) + t(1-t)\phi(\|x-y\|) \leq tf(x) + (1-t)f(y) \ \forall x, y \in \operatorname{dom} f \text{ and } \forall t \in (0,1).$$

In case this inequality holds for $\phi = (\beta/2)(\cdot)^2$, where $\beta > 0$, then f is said to be β-*strongly convex*. Let us mention that this property implies β-strong monotonicity of ∂f (see [6, Example 22.3]) (more general, if f is uniformly convex, then ∂f is uniformly monotone, see [6, Example 22.3]).

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

$$J_{\gamma N_S} = J_{N_S} = J_{\partial \delta_S} = (\operatorname{Id}_{\mathcal{H}} + N_S)^{-1} = \operatorname{prox}_{\delta_S} = P_S, \tag{37}$$

where $P_S : \mathcal{H} \to C$ denotes the *orthogonal projection operator* on S (see [6, Example 23.3 and Example 23.4]).

In the sequel we consider the following primal-dual pair of convex optimization problems.

Problem 14 Let $\mathcal{H}$ be a real Hilbert space and let $f \in \Gamma(\mathcal{H})$, $z \in \mathcal{H}$. Let m be a strictly positive integer and for every $i \in \{1,...,m\}$, suppose that $\mathcal{G}_i$ is a real Hilbert space, let $g_i, l_i \in \Gamma(\mathcal{G}_i)$, $r_i \in \mathcal{G}_i$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero bounded linear operator. Consider the convex optimization problem

$$(P) \quad \inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \square l_i)(L_i x - r_i) - \langle x, z \rangle \right\} \tag{38}$$

and its conjugate dual problem

$$(D) \quad \sup_{(v_1,...,v_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m} \left\{ -f^* \left(z - \sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} (g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i \rangle) \right\}. \tag{39}$$

By taking into account the maximal monotone operators

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

the monotone inclusion problem (22) reads

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in \partial f(\overline{x}) + \sum_{i=1}^{m} L_i^*(\partial g_i \square \partial l_i)(L_i \overline{x} - r_i), \tag{40}$$

while the dual inclusion problem (23) reads

$$\text{find } \overline{v}_1 \in \mathcal{G}_1, \ ..., \overline{v}_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in \partial f(x) \\ \overline{v}_i \in (\partial g_i \square \partial l_i)(L_i x - r_i), \ i = 1,...,m. \end{cases} \tag{41}$$

If $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 ... \times \mathcal{G}_m$ is a primal-dual solution to (40)–(41), namely,

$$z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in \partial f(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \square \partial l_i)(L_i \overline{x} - r_i), \ i = 1,...,m, \tag{42}$$

15

then $\bar{x}$ is an optimal solution to (P), $(\bar{v}_1,...,\bar{v}_m)$ 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).

Combining this statement with Algorithm 12 and Theorem 13 gives rise to the following iterative scheme and corresponding convergence theorem for the primal-dual pair of optimization problems (P)–(D).

Algorithm 15 *Let $x_0, x_1 \in \mathcal{H}$, $v_{i,0}, v_{i,1} \in \mathcal{G}_i$, $i = 1,...,m$, and $\tau, \sigma_i > 0$, $i = 1,...,m$, be such that*

$$\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < 4.$$

Furthermore, let $(\alpha_n)_{n \geq 1}$ be a nondecreasing sequence with $\alpha_1 = 0$ and $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ and the sequence $(\lambda_n)_{n \geq 1}$ be such that

$$\delta > \frac{\alpha^2(1+\alpha) + \alpha\sigma}{1-\alpha^2} \quad and \quad 0 < \lambda \leq \lambda_n \leq 2\frac{\delta - \alpha\left[\alpha(1+\alpha) + \alpha\delta + \sigma\right]}{\delta\left[1 + \alpha(1+\alpha) + \alpha\delta + \sigma\right]} \quad \forall n \geq 1.$$

Set

$$(\forall n \geq 1) \left|
\begin{array}{l}
p_{1,n} = \mathrm{prox}_{\tau f}\big(x_n + \alpha_n(x_n - x_{n-1}) - \frac{\tau}{2}\sum_{i=1}^{m} L_i^*(v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1})) + \tau z\big) \\
w_{1,n} = 2p_{1,n} - x_n - \alpha_n(x_n - x_{n-1}) \\
\text{For } i = 1,...,m \\
\quad \left\lfloor \begin{array}{l}
p_{2,i,n} = \mathrm{prox}_{\sigma_i g_i^*}\big(v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1}) + \frac{\sigma_i}{2}L_i w_{1,n} - \sigma_i r_i\big) \\
w_{2,i,n} = 2p_{2,i,n} - v_{i,n} - \alpha_n(v_{i,n} - v_{i,n-1})
\end{array} \right. \\
z_{1,n} = w_{1,n} - \frac{\tau}{2}\sum_{i=1}^{m} L_i^* w_{2,i,n} \\
x_{n+1} = x_n + \alpha_n(x_n - x_{n-1}) + \lambda_n(z_{1,n} - p_{1,n}) \\
\text{For } i = 1,...,m \\
\quad \left\lfloor \begin{array}{l}
z_{2,i,n} = \mathrm{prox}_{\sigma_i l_i^*}\big(w_{2,i,n} + \frac{\sigma_i}{2}L_i(2z_{1,n} - w_{1,n})\big) \\
v_{i,n+1} = v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1}) + \lambda_n(z_{2,i,n} - p_{2,i,n}).
\end{array} \right.
\end{array} \right.$$

(43)

Theorem 16 *In Problem 14, suppose that*

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

(44)

and consider the sequences generated by Algorithm 15. Then there exists $(\bar{x}, \bar{v}_1,..., \bar{v}_m) \in \mathcal{H} \times \mathcal{G}_1 ... \times \mathcal{G}_m$ such that the following statements are true:

(i) By setting

$$\bar{p}_1 = \mathrm{prox}_{\tau f}\left(\bar{x} - \frac{\tau}{2}\sum_{i=1}^{m} L_i^* \bar{v}_i + \tau z\right),$$

$$\bar{p}_{2,i} = \mathrm{prox}_{\sigma_i g_i^*}\left(\bar{v}_i + \frac{\sigma_i}{2}L_i(2\bar{p}_1 - \bar{x}) - \sigma_i r_i\right), \quad i = 1,...,m,$$

the element $(\bar{p}_1, \bar{p}_{2,1},..., \bar{p}_{2,m}) \in \mathcal{H} \times \mathcal{G}_1 \times... \times \mathcal{G}_m$ is a primal-dual solution to Problem 14, hence $\bar{p}_1$ is an optimal solution to (P) and $(\bar{p}_{2,1},..., \bar{p}_{2,m})$ is an optimal solution to (D);

(ii) $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$, and $\sum_{n\in\mathbb{N}} \|v_{i,n+1} - v_{i,n}\|^2 < +\infty$, $i = 1,...,m$;

(iii) $(x_n, v_{1,n},..., v_{m,n})_{n\in\mathbb{N}}$ converges weakly to $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$;

(iv) $(p_{1,n} - z_{1,n}, p_{2,1,n} - z_{2,1,n},..., p_{2,m,n} - z_{2,m,n}) \to 0$ as $n \to +\infty$;

(v) $(p_{1,n}, p_{2,1,n},..., p_{2,m,n})_{n\geq1}$ converges weakly to $(\overline{p}_1, \overline{p}_{2,1},..., \overline{p}_{2,m})$;

(vi) $(z_{1,n}, z_{2,1,n},..., z_{2,m,n})_{n\geq1}$ converges weakly to $(\overline{p}_1, \overline{p}_{2,1},..., \overline{p}_{2,m})$;

(vii) if f and g_i^*, $i = 1,...,m$, are uniformly convex, then $(p_{1,n}, p_{2,1,n},..., p_{2,m,n})_{n\geq1}$ and $(z_{1,n}, z_{2,1,n},..., z_{2,m,n})_{n\geq1}$ converge strongly to the unique primal-dual solution $(\overline{p}_1, \overline{p}_{2,1},..., \overline{p}_{2,m})$ to Problem 14.

We refer the reader to [15, 23] for qualification conditions which guarantee that the inclusion in (44) holds. Finally, let us mention that for $i = 1,...,m$, the function g_i^* is uniformly convex if it is α_i-strongly convex for $\alpha_i > 0$ and this is the case if and only if g_i is Fréchet-differentiable with α_i^{-1}-Lipschitz gradient (see [6, Theorem 18.15]).

5 Numerical experiments

5.1 Clustering

In cluster analysis one aims for grouping a set of points such that points in the same group are more similar to each other than to points in other groups. Let $u_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, be given points. For each point u_i we are looking for determining the associated cluster center $x_i \in \mathbb{R}^n$, $i = 1, \ldots, m$. By taking into account [21, 27], clustering can be formulated as the convex optimization problem

$$\inf_{x_i \in \mathbb{R}^n, i=1,...,m} \left\{ \frac{1}{2} \sum_{i=1}^m \|x_i - u_i\|^2 + \gamma \sum_{i<j} \omega_{ij} \|x_i - x_j\|_p \right\}, \qquad (45)$$

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 $i, j \in \{1, \ldots, m\}$, $i < j$. Since the objective function is strongly convex, there exists a unique solution to (45).

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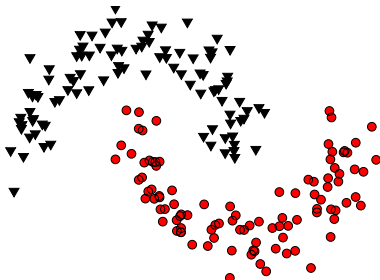

Figure 1: Clustering two interlocking half moons. The colors (resp. the shapes) show the correct affiliations.

	$p = 2,\ \gamma = 5.2$		$p = 1,\ \gamma = 4$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$
Algorithm 15	0.65s (175)	1.36s (371)	0.63s (176)	1.27s (374)
DR [15]	0.78s (216)	1.68s (460)	0.78s (218)	1.68s (464)
FB [41]	2.48s (1353)	5.72s (3090)	2.01s (1092)	4.05s (2226)
FB Acc [13]	2.04s (1102)	4.11s (2205)	1.74s (950)	3.84s (2005)
FBF [23]	7.67s (2123)	17.58s (4879)	6.33s (1781)	13.22s (3716)
FBF Acc [14]	5.05s (1384)	10.27s (2801)	4.83s (1334)	9.98s (2765)
PD [18]	1.48s (780)	3.26s (1708)	1.44s (772)	3.18s (1722)
PD Acc [18]	1.28s (671)	3.14s (1649)	1.23s (665)	3.12s (1641)
Nesterov [34]	7.85s (3811)	42.69s (21805)	7.46s (3936)	> 190s (> 100000)
FISTA [7]	7.55s (4055)	51.01s (27356)	6.55s (3550)	47.81s (26069)

Table 1: 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. The tuning parameter γ is chosen in order to guarantee a correct separation of the input data into the two half moons.

In addition to this, the choice of the weights is important as well, since cluster centers may coalesce immediately as γ passes certain critical values. In terms of our weight selection, we use a K-nearest neighbors strategy, as proposed in [21]. Therefore, whenever $i, j \in \{1, \ldots, m\}$, $i < j$, 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 use the values $K = 10$ and $\phi = 0.5$, which are the best ones reported in [21] on a similar dataset.

Let k be the number of nonzero weights ω_{ij}. Then, one can introduce a linear operator $A : \mathbb{R}^{mn} \to \mathbb{R}^{kn}$, such that problem (45) can be equivalently written as

$$\inf_{x \in \mathbb{R}^{mn}} \{h(x) + g(Ax)\}, \tag{46}$$

the function h being 1-strongly convex and differentiable with 1-Lipschitz continuous gradient. Also, by taking $p \in \{1, 2\}$, the proximal points with respect to g^* are known to be available via explicit formulae.

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

Given Table 1, it shows that Algorithm 15 performs better than the noninertial Douglas–Rachford (DR) method proposed in [15, Algorithm 2.1]. One can also see that the inertial Douglas–Rachford algorithm is faster than other popular primal-dual solvers, among them the forward-backward-forward (FBF) method from [23], and the forward-backward (FB) method from [41], where in both methods the function h is processed via a

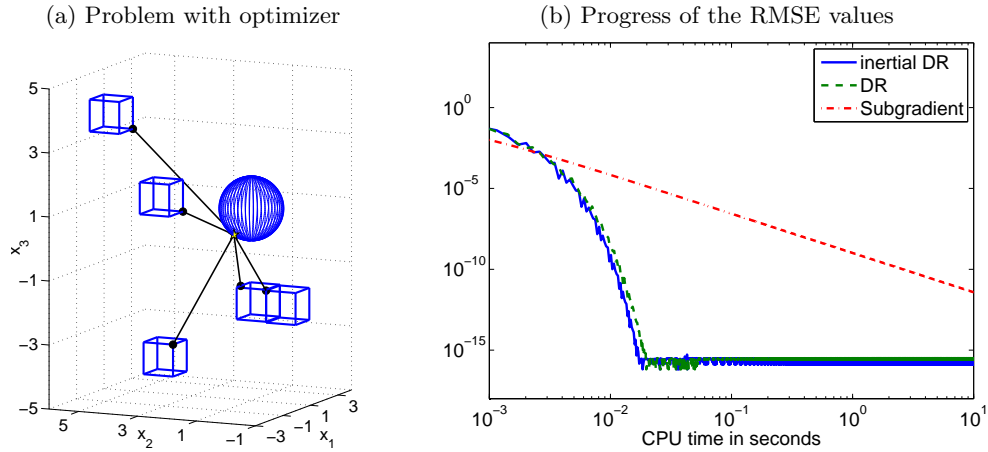(a) Problem with optimizer   (b) Progress of the RMSE values

Figure 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.

forward step. The accelerated versions of the latter and of the primal-dual (PD) method from [18] converge in less time than their regular variants, but are still slower than Algorithm 12. Notice that the methods called Nesterov and FISTA are accelerated proximal gradient algorithms which are applied to the Fenchel dual problem to (46).

5.2   The generalized Heron problem

In the sequel we investigate the *generalized Heron problem* which has been recently investigated in [31, 32] and where for its solving subgradient-type methods have been proposed.

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 its distances to two given points is minimal, the problem that we address here aims to find a point in a closed convex set $\Omega \subseteq \mathbb{R}^n$ which minimizes the sum of the distances to given convex closed sets $\Omega_i \subseteq \mathbb{R}^n$, $i = 1, \ldots, m$.

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

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

Thus the *generalized Heron problem* reads

$$\inf_{x \in \Omega} \sum_{i=1}^m d(x; \Omega_i), \tag{47}$$

where the sets $\Omega \subseteq \mathbb{R}^n$ and $\Omega_i \subseteq \mathbb{R}^n$, $i = 1, \ldots, m$, are assumed to be nonempty, closed and convex. We observe that (47) perfectly fits into the framework considered in Problem 14 when setting

$$f = \delta_\Omega, \text{ and } g_i = \| \cdot \|, \ l_i = \delta_{\Omega_i} \text{ for all } i = 1, \ldots, m. \tag{48}$$

However, note that (47) cannot be solved via the primal-dual methods in [23] and [41], which require for each $i = 1, ..., m$, that either g_i or l_i is strongly convex, unless one substantially increases the number of primal and dual variables. Notice that

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

19

	Algorithm 15		Douglas–Rachford, [15]		Subgradient, [31, 32]	
	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-10}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-10}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-10}$
$n = 2, \; m = 5$	0.01s (33)	0.03s (72)	0.01s (30)	0.03s (63)	–	–
$n = 2, \; m = 10$	0.01s (21)	0.03s (59)	0.01s (21)	0.02s (43)	0.01s (8)	0.03s (120)
$n = 2, \; m = 20$	0.06s (295)	0.11s (522)	0.11s (329)	0.19s (583)	0.05s (204)	16.78s (69016)
$n = 2, \; m = 50$	0.18s (517)	0.45s (1308)	0.22s (579)	0.55s (1460)	0.04s (152)	4.82s (19401)
$n = 3, \; m = 5$	0.01s (16)	0.01s (37)	0.01s (16)	0.01s (33)	0.02s (70)	2.17s (8081)
$n = 3, \; m = 10$	0.01s (37)	0.03s (91)	0.01s (41)	0.03s (101)	0.01s (11)	0.03s (199)
$n = 3, \; m = 20$	0.01s (22)	0.03s (52)	0.01s (25)	0.03s (59)	0.01s (6)	0.01s (32)
$n = 3, \; m = 50$	0.01s (19)	0.02s (44)	0.01s (21)	0.02s (51)	0.01s (10)	0.01s (17)

Table 2: Performance evaluation for the Heron problem. The entries refer to the CPU times in seconds and the number of iterations, respectively, needed in order to attain a root-mean-square error lower than the tolerance ε.

where $\overline{B}(0, 1)$ denotes the closed unit ball, thus the proximal points of f, g_i^* and l_i^*, $i = 1, \ldots, m$, can be calculated via projections, in case of the latter via Moreau's decomposition formula (36).

In the following we solve a number of random problems where the closed convex set $\Omega \subseteq \mathbb{R}^n$ will always be the unit ball centered at $(1,...,1)^T$. The sets $\Omega_i \subseteq \mathbb{R}^n$, $i = 1,...,m$, are boxes in right position (i.e., the edges are parallel to the axes) with side length 1. The box centers are created via independent identically distributed Gaussian entries from $\mathcal{N}(0, n^2)$ where the random seed in Matlab is set to 0. After determining a solution, the stopping criterion asks the root-mean-square error (RMSE) to be less than or equal to a given bound ε.

Table 2 shows a comparison between Algorithm 15, the Douglas–Rachford type method from [15, Algorithm 3.1], and the subgradient approach described in [31, 32] when applied to different instances of the generalized Heron problem. One such particular case is displayed in Figure 2 when $n = 3$ and $m = 5$, while the evolution of the RMSE values is given there in more detail. Empty cells in Table 2 indicate that it took more than 60 seconds to pass the stopping criterion. Based on the provided data, one can say that both Algorithm 15 and the noninertial Douglas–Rachford type method are performing well in this example and that differences in the computational performance are almost negligible. However, one very interesting observation arises when the dimension of the space is set to $n = 3$, as the subgradient approach then becomes better and surpasses both primal-dual methods.

References

[1] F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM Journal on Control and Optimization 38(4), 1102–1119, 2000

[2] F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM Journal on Optimization 14(3), 773–782, 2004

[3] F. Alvarez, H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis 9, 3–11, 2001

[4] H. Attouch, J. Peypouquet, P. Redont, *A dynamical approach to an inertial forward-backward algorithm for convex minimization*, SIAM Journal on Optimization 24(1), 232–256, 2014

[5] H. Attouch, M. Théra, *A general duality principle for the sum of two operators*, Journal of Convex Analysis 3, 1–24, 1996

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

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

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

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

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

[11] R.I. Boţ, E.R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, arXiv:1402.5291, 2014

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

[13] R.I. Boţ, E.R. Csetnek, A. Heinrich, C. Hendrich, *On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems*, Mathematical Programming, DOI 10.1007/s10107-014-0766-0

[14] R.I. Boţ, C. Hendrich, *Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization*, Journal of Mathematical Imaging and Vision, 49(3), 551–568, 2014

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

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

[17] A. Cabot, P. Frankel, *Asymptotics for some proximal-like method involving inertia and memory aspects*, Set-Valued and Variational Analysis 19, 59–74, 2011

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

[19] R.H. Chan, S. MA, J. Yang, *Inertial primal-dual algorithms for structured convex optimization*, arXiv:1409.2992v1, 2014

[20] C. Chen, S. MA, J. Yang, *A general inertial proximal point method for mixed variational inequality problem*, arXiv:1407.8238v2, 2014

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

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

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

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

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

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

[27] F. Lindsten, H. Ohlsson, L. Ljung, *Just relax and come clustering! A convexication of k-means clustering*, Technical Report, Linköpings universitet, 2011

[28] P.L. Lions, B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis 16(6), 964–979, 1979

[29] P.-E. Maingé, *Convergence theorems for inertial KM-type algorithms*, Journal of Computational and Applied Mathematics 219, 223–236, 2008

[30] P.-E. Maingé, A. Moudafi, *Convergence of new inertial proximal methods for dc programming*, SIAM Journal on Optimization 19(1), 397–413, 2008

[31] B.S. Mordukhovich, N.M. Nam and J. Salinas, *Solving a generalized Heron problem by means of convex analysis*, American Mathematical Monthly 119(2), 87–99, 2012

[32] B.S. Mordukhovich, N.M. Nam and J. Salinas, *Applications of variational analysis to a generalized Heron problem*, Applicable Analysis 91(10), 1915–1942, 2012

[33] A. Moudafi, M. Oliny, *Convergence of a splitting inertial proximal method for monotone operators*, Journal of Computational and Applied Mathematics 155, 447–454, 2003

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

[35] P. Ochs, Y. Chen, T. Brox, T. Pock, *iPiano: Inertial proximal algorithm for non-convex optimization*, SIAM Journal on Imaging Sciences 7(2), 1388–1419, 2014

[36] J.-C. Pesquet, N. Pustelnik, *A parallel inertial proximal optimization method*, Pacific Journal of Optimization 8(2), 273–306, 2012

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

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

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

[40] B. F. Svaiter, *On weak convergence of the Douglas-Rachford method*, SIAM Journal on Control and Optimization 49(1), 280–287, 2011

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

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