Recent advances in splitting methods for monotone inclusions and nonsmooth optimization problems

Ernö Robert Csetnek

Habilitation Thesis
(defended on 2 March 2016)

Faculty of Mathematics

University of Vienna

Preface

Several results concerning the solving of monotone inclusion problems by splitting methods obtained by the author in the last years are presented in this thesis. For convex optimization problems, the optimality conditions characterizing the set of solutions can be written in many cases in the form of monotone inclusion problems, justifying the study of problems where the sum of maximally monotone operators is involved. The present thesis aims to outline the most important contributions of the author to theoretical results and some of their algorithmic realizations in convex (nondifferentiable) optimization and monotone operator theory.

The first part of the thesis addresses the rate of convergence of a primal-dual splitting method for solving highly structured monotone inclusion problems. Primal-dual algorithms of proximal-type are numerical schemes that solve efficiently primal-dual pairs of monotone inclusions and convex optimization problems consisting of sums, linear compositions, parallel sums, and infimal convolutions by making use of the resolvents of the monotone operators involved. They are fully decomposable in the sense that each operator is evaluated in the algorithm separately.

Proximal-point type algorithms with inertial and memory effects are also addressed. The incorporation of inertial terms in splitting algorithms is motivated by the discretization of a differential system of second-order in time, called heavy-ball method. We focus our attention on the inertial versions of the forward-backward-forward and Douglas-Rachford splitting methods. Furthermore, we investigate an inertial proximal-type splitting method for nonconvex optimization problems.

We consider penalty-type splitting algorithms for variational inequalities written as monotone inclusion problems. We investigate a forward-backward and a Tseng's type numerical scheme, the latter allowing us to formulate penalty-type splitting algorithms for even more complicated monotone inclusion problems involving finite sums and compositions with linear operators. In particular, we are able to solve convex optimization problems with intricate objective functions over the set of minima of a convex and differentiable function.

In the last part of the thesis we approach the solving of monotone inclusion problems via first and second order dynamical systems of implicit-type. These are ordinary differential equations formulated by making use of the resolvents of the monotone operators involved. The existence of the trajectories is guaranteed in the framework of the Cauchy-Lipschitz-Picard Theorem, while the (weak) asymptotical convergence of the orbits to a solution is based on Lyapunov analysis. In the asymptotic analysis performed we report also several results concerning the rate of convergence of the trajectories and, in some cases, of the objective functions along the orbits, when considering convex optimization problems.

Acknowledgements. I want to use this opportunity to thank Radu Ioan Boţ for the excellent scientific collaboration which started at the Chemnitz University of Technology in 2005, when I began my PhD studies under his supervisions, and which continued at the University of Vienna. His guidance helped me in all the significant steps in my academic career. It was a wonderful journey through many interesting and challenging research fields, starting with theoretical aspects in convex analysis and monotone operator theory and arriving also to real-life problems. Thanks also for his advice and friendliness.

I would like to thank also Gert Wanka, the head of the research group Approximation Theory at the Faculty of Mathematics of the Chemnitz University of Technology, for the support during these years. Many thanks also to Sorin-Mihai Grad, André Heinrich and Christopher Hendrich, former and current members of the aforementioned group and to Sebastian Banert for stimulating discussions concerning several research topics.

I am grateful to the Department of Mathematics, University of Vienna, for providing me an excellent research environment.

I acknowledge financial support from DFG (German Research Foundation) and FWF (Austrian Science Fund) via the projects BO 2516/4-1 and M 1682-N25, respectively.

Many thanks to my parents for understanding and support.

Last, but not least, I want especially to thank my wife Minodora and my daughter Hannah for their unconditioned love, patience and encouragements.

Ernö Robert Csetnek
September 28, 2015
Vienna
Austria

Contents

Chapter 1

Introduction

The aim of this work is to present a number of contributions in the context of solving monotone inclusion problems in Hilbert spaces by means of algorithmic schemes of proximal-splitting-type, a setting which allows the numerical treatment of highly structured nondifferentiable convex optimization problems with intricate objective function. Due to its numerous applications in signal and image processing, portfolio optimization, cluster analysis, location theory, average consensus on colored networks, image classification via support vector machines (and this enumeration of fields can be continued), this topic is in the last couple of years of huge interest for the applied mathematics community.

Finding the set of zeros of monotone operators is motivated by the fact that optimality conditions for convex optimization problems which fulfill a regularity condition can be expressed as monotone inclusion problems. Furthermore, the investigations performed in this more general setting of (maximally) monotone operators bring new insights when considering the problem of solving complicated nondifferentiable convex optimization problems involving finite sums, compositions with linear operators or infimal convolutions. Moreover, due to its applications in the theory of nonlinear partial differential equations, variational inequalities and optimization theory, the study of monotone inclusions continuous to attract many mathematicians.

Let us briefly recall the fundamental proximal-splitting algorithms from the literature in their simpler form, namely the proximal-point algorithm, the forward-backward splitting, the forward-backward-forward scheme and the Douglas-Rachford splitting, respectively. One of the first algorithms of this type has been proposed and analysed by Rockafellar [122] in connection with the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax, \tag{1. 1}$$

where $\mathcal{H}$ is a real Hilbert space and $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator. The so-called *proximal-point* algorithm generates iteratively a sequence as follows: chose $x_0 \in \mathcal{H}$ and for $n \geq 0$ set

$$x_{n+1} = J_{\eta A}(x_n), \tag{1. 2}$$

where $\eta > 0$ and $J_A : \mathcal{H} \to \mathcal{H}$, defined by $J_A = (\mathrm{Id}_{\mathcal{H}} + A)^{-1}$, is the resolvent of A. The properties of the latter rely on a seminal result due to Minty in Hilbert spaces, saying that the sum of the identity and a maximally monotone operator is surjective. Let us underline that, when considering numerical schemes of this type, the usual terminology is that we perform a *backward* step, meaning that the set-valued operator is evaluated via its resolvent. The asymptotic analysis of the above algorithm reveals that the sequence generated by (1. 2) weakly converges to a solution of (1. 1), provided the set of solutions of the latter is nonempty.

Assume now that one is interested in solving the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx, \tag{1. 3}$$

where $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ are maximally monotone operators. Since in general there exists no closed formula for the resolvent of the sum of two operators in terms of their resolvents, the above algorithm (1. 2) is from implementation point of view not suitable for solving (1. 3). The so-called *splitting* algorithms overcome this drawback, where the word "splitting" is used in order to stress out that in the iterative schemes the operators involved are evaluated separately.

For the beginning assume that $B : \mathcal{H} \to \mathcal{H}$ is a (single-valued) β-cocoercive operator, for $\beta > 0$. The *forward-backward* algorithm has the following form (see for example [26]): chose $x_0 \in \mathcal{H}$ and for $n \geq 0$ set

$$x_{n+1} = J_{\eta A}(x_n - \eta B x_n), \tag{1. 4}$$

where $\eta \in (0, 2\beta)$. In this case the sequence generated by (1. 4) converges weakly to a solution of (1. 3), as soon as the set of solutions of the latter is nonempty. The terminology forward-backward is justified by the fact that the set-valued mapping is evaluated through a backward step and the single-valued one via a *forward* step.

Let us suppose now that the cocoercivity of B is relaxed to monotonicity and Lipschitz-continuity. Under these premises, Tseng's numerical scheme [128, 129], also called *forward-backward-forward* algorithm, solves the problem (1. 3) according to (see also [26, 62]): chose $x_0 \in \mathcal{H}$ and for $n \geq 0$ set

$$p_n = J_{\lambda_n A}(x_n - \lambda_n B x_n) \tag{1. 5}$$
$$x_{n+1} = p_n + \lambda_n (B x_n - B p_n), \tag{1. 6}$$

where for all $n \geq 0$, $\lambda_n \in [\varepsilon, (1-\varepsilon)/\beta]$ with $\varepsilon \in (0, 1/(\beta+1))$, β being the Lipschitz parameter of B. If we assume that the set of solutions to (1. 3) is nonempty, both sequences generated by (1. 5)-(1. 6) converge to a solution of the monotone inclusion problem. Despite the fact that the forward-backward-forward algorithm requires an additional sequence to be computed, it turned out that this numerical scheme opens the gate towards the development of the so-called primal-dual algorithms that are able to solve highly structured monotone inclusion problems (see [62]).

Finally, in case in (1. 3) both of the operators A and B are set-valued, the *Douglas-Rachford* algorithm (see [26, 83, 97]) solves (1. 3) by the numerical scheme: chose $x_0 \in \mathcal{H}$ and for $n \geq 0$ set

$$y_n = J_{\eta B}(x_n), \tag{1. 7}$$
$$z_n = J_{\eta A}(2y_n - x_n), \tag{1. 8}$$
$$x_{n+1} = x_n + z_n - y_n, \tag{1. 9}$$

where $\eta > 0$ is arbitrary chosen. If the set of solutions to (1. 3) is nonempty, then the sequences $(y_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge weakly to the same solution of (1. 3). We refer also to [84] for further investigations on the Douglas-Rachford algorithm, where it has been pointed out that this numerical scheme can be viewed as a proximal-point algorithm (1. 2) for a particular maximal monotone operator.

In the following we will present a short historical overview of those further developments of the proximal methods which are relevant to this thesis, namely, of the primal-dual proximal splitting methods, the inertial-type algorithms, the penalty-type numerical schemes and the dynamical systems of implicit-type associated to monotone inclusion problems.

We come now to more involved monotone inclusion problems. Whenever one of the operators in (1. 3) is replaced by the composition of a maximally monotone operator with a linear and continuous mapping, one faces major difficulties in applying the aforementioned splitting methods, since the resolvent of such a composition cannot be expressed in a closed form (excepting some very restrictive cases). The modern techniques called *primal-dual* methods overcome this difficulty, see [59, 62, 76, 130]. First results concerning proximal-type splitting algorithms for solving convex optimization problems where compositions with linear and continuous operators are involved have been reported by Combettes and Ways [78], Esser, Zhang and Chan [86] and Chambolle and Pock [69].

Further investigations in the framework of monotone inclusion problems have been performed by Briceño-Arias and Combettes [62]. They treated monotone inclusion problems involving sums of compositions with linear and continuous operators by rewriting the original monotone inclusion problem as the sum of a maximally monotone operator and a linear and skew one in an appropriate product space, which has been solved through the aforementioned Tseng's algorithm (1. 5)-(1. 6). We refer the reader to [62] and [76] for this forward-backward-forward primal-dual splitting algorithm. Moreover, by taking advantage again of the product space approach, this time in a suitable renormed space, Vũ succeeded in [130] to give a primal-dual splitting algorithm of forward-backward type. Finally, by using some techniques from [130], Boţ and Hendrich presented in [59] a primal-dual algorithm of Douglas-Rachford type.

Let us underline some highlights of the primal-dual splitting algorithms. These are methods which solve concomitantly a primal inclusion problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + \sum_{i=1}^{m} L_i^*\big((B_i \Box D_i)(L_i x)\big) + Cx \tag{1. 10}$$

together with the dual inclusion in the sense of Attouch-Théra [21]:

$$\text{find } v_1 \in \mathcal{G}_1, \ldots, v_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \left\{ \begin{array}{l} -\sum_{i=1}^{m} L_i^* v_i \in Ax + Cx \\ v_i \in (B_i \Box D_i)(L_i x),\ i = 1, \ldots, m, \end{array} \right. \tag{1. 11}$$

where $\mathcal{H}$ and $\mathcal{G}_i$, $i = 1, ..., m$ are real Hilbert spaces, $A : \mathcal{H} \rightrightarrows \mathcal{H}$, $C : \mathcal{H} \to \mathcal{H}$, $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ are maximally monotone operators and $L_i : \mathcal{H} \to \mathcal{G}_i$ are nonzero linear continuous operators, $i = 1, ..., m$.

The primal-dual splitting algorithms are fully decomposable, in the sense that each operator is evaluated separately in the iterative scheme. The splitting schemes can be used for solving highly structured nondifferentiable convex optimization problems of the from

$$\inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \Box l_i)(L_i x) + h(x) \right\} \tag{1. 12}$$

and their *Fenchel-type dual* problems

$$\sup_{v_i \in \mathcal{G}_i,\, i=1,..,m} \left\{ -\big(f^* \Box h^*\big) \left(-\sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} \big(g_i^*(v_i) + l_i^*(v_i) \big) \right\} \tag{1. 13}$$

expressed by means of the Fenchel conjugates of the functions involved.

Considering compositions with linear and continuous operators is motivated by the fact that in image processing the discrete first order total variational functional used in the reconstruction of images can be represented as such a composition. The use of infimal convolutions is justified by the fact that the second order total variational functional can be expressed in this way, see [29, 57]. Moreover, infimal convolutions appear naturally also when solving generalized location problems like the Heron problem, which aims to find a point in a closed convex set which minimizes the sum of distances to given convex closed sets.

Let us come now to the class of so-called *inertial proximal methods*. The idea behind these iterative schemes relies on the use of an implicit discretization of a differential system of second-order in time, called *heavy ball method*. 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. This advantage of taking into account the "prehistory" of the process could accelerate the convergence of the iterates, as observed for example by Polyak [118] in case of minimizing differentiable functions. Let us mention here also the *fast gradient method* of Nesterov [105] and the so-called FISTA (see [28]), which are iterative schemes involving some inertial terms which are able to accelerate the convergence for the objective function values.

As emphasized by Ochs, Chen, Brox and Pock in [110, Section 5.1] and Bertsekas in [30, Exercise 1.3.9] one of the aspects which makes algorithms with inertial/memory effects useful is the fact that they are able to detect local optimal solutions of (nonconvex) minimization problems which cannot be found by their non-inertial variants.

Inertial-type algorithms have been considered for the first time in the context of monotone inclusion problems by Alvarez and Attouch in [3,5]. The iterative scheme proposed in [5] (in its simplified version) reads as: chose $x_0, x_1 \in \mathcal{H}$ and for $n \geq 1$ set

$$x_{n+1} = J_{\eta A}(x_n + \alpha(x_n - x_{n-1})). \tag{1. 14}$$

Under appropriate conditions imposed on the step size $\eta > 0$ and on the parameter $\alpha \geq 0$ controlling the inertial term, the generated sequence of iterates converges weakly to a solution of (1. 1).

Especially noticeable is that these ideas have been also used in the context of the problem (1. 3) in case B is a (single-valued) cocoercive operator, giving rise to the so-called inertial forward-backward algorithm considered by Moudafi and Oliny in [104]: chose $x_0, x_1 \in \mathcal{H}$ and for $n \geq 1$ set

$$x_{n+1} = J_{\eta A}(x_n - \eta B x_n + \alpha(x_n - x_{n-1})). \tag{1. 15}$$

One can notice a considerable interest in the class of inertial type algorithms, see also the works of Alvarez [4], Cabot and Frankel [67], Maingé [99], [100], Pesquet and Pustelnik [116]. We mention here also the works of Chen, Chan, Ma and Yang [70,71] and Ghadimi, Feyzmahdavian and Johansson [89], where further convergence rates for several inertial type algorithms have been reported.

We turn now our attention to *penalty-type* proximal splitting methods. These are designed to solve variational inequalities expressed as monotone inclusion problems of the form

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + N_M(x), \tag{1. 16}$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator, $M = \operatorname{argmin} \Psi$ is the set of global minima of the convex function $\Psi : \mathcal{H} \to \mathbb{R}$, which is supposed to be differentiable with Lipschitz continuous gradient fulfilling $\min \Psi = 0$, and $N_M : \mathcal{H} \rightrightarrows \mathcal{H}$ is the normal cone of the set $M \subseteq \mathcal{H}$ (see the works of Attouch, Czarnecki and Peypouquet [15,16], Noun and Peypouquet [109,115]. Specifically, one can find in the literature forward-backward-type algorithms for solving (4. 3) (see [15, 16, 109, 115]), which perform in each iteration a proximal step with respect to A and a gradient step with respect to the penalization function Ψ: chose $x_1 \in \mathcal{H}$ and for $n \geq 1$ set

$$x_{n+1} = J_{\lambda_n A}(x_n - \lambda_n \beta_n \nabla \Psi(x_n)), \tag{1. 17}$$

with $(\lambda_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ sequences of positive real numbers. Ergodic convergence results are usually obtained assuming the fulfillment of a conditions expressed by means of the conjugate function of Ψ, which is the discretized counterpart of a condition introduced by Attouch and Czarnecki in [14] in the context of continuous-time nonautonomous differential inclusions.

It is worth mentioning that when A is the convex subdifferential of a proper, convex and lower semicontinuous function $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, the above algorithm provides an iterative scheme for solving convex optimization problems which can be formulated as

$$\min_{x \in \mathcal{H}} \{\Phi(x) : x \in \operatorname{argmin} \Psi\}. \tag{1. 18}$$

Finally, let us return to (1. 3) and say a few words about continuous *implicit-type dynamical systems* associated with this problem, which are ordinary differential equations formulated via resolvents of maximal monotone operators. In [31], Bolte studied the convergence of the trajectories of the following dynamical system

$$\begin{cases} \dot{x}(t) + x(t) = \operatorname{proj}_C \big(x(t) - \eta \nabla \phi(x(t))\big) \\ x(0) = x_0. \end{cases} \tag{1. 19}$$

where $\phi : \mathcal{H} \to \mathbb{R}$ is a convex C^1 function defined on a real Hilbert space $\mathcal{H}$, C is a nonempty, closed and convex subset of $\mathcal{H}$, $x_0 \in \mathcal{H}$, $\eta > 0$ and proj_C denotes the projection operator on the set C. In this context it is shown that the trajectory of (1. 19) converges weakly to a minimizer of the optimization problem

$$\inf_{x \in C} \phi(x), \tag{1. 20}$$

provided the latter is solvable. We refer also to the work of Antipin [7] for further statements and results concerning (1. 19).

The following generalization of the dynamical system (1. 19) has been recently considered by Abbas and Attouch in [1, Section 5.2]:

$$\begin{cases} \dot{x}(t) + x(t) = \operatorname{prox}_{\eta\Phi} \big(x(t) - \eta B(x(t))\big) \\ x(0) = x_0, \end{cases} \tag{1. 21}$$

where $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function defined on a real Hilbert space $\mathcal{H}$, $B : \mathcal{H} \to \mathcal{H}$ is an η-cocoercive operator, $x_0 \in \mathcal{H}$, $\eta > 0$ and $\operatorname{prox}_{\eta\Phi} : \mathcal{H} \to \mathcal{H}$,

$$\operatorname{prox}_{\eta\Phi}(x) = \operatorname*{argmin}_{y \in \mathcal{H}} \left\{ \Phi(y) + \frac{1}{2\eta}\|y - x\|^2 \right\}, \tag{1. 22}$$

denotes the proximal point operator of $\eta\Phi$.

According to [1], in case $\operatorname{zer}(\partial\Phi + B) \neq \emptyset$, the weak asymptotical convergence of the orbit x of (1. 21) to an element in $\operatorname{zer}(\partial\Phi + B) \neq \emptyset$ is ensured by choosing the step-size η in a suitable domain bounded by the parameter of cocoercivity of the operator B (notice that $\partial\Phi$ denotes the convex subdifferential of Φ and $\operatorname{prox}_{\eta\Phi} = J_{\eta\partial\Phi}$).

For the minimization of the smooth and convex function $g : \mathcal{H} \to \mathbb{R}$ over the nonempty, convex and closed set $C \subseteq \mathcal{H}$, a continuous in time second order gradient-projection approach has been considered in [7, 8], having as starting point the dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) = \operatorname{proj}_C(x(t) - \eta\nabla g(x(t))) \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{1. 23}$$

with constant damping parameter $\gamma > 0$ and constant step size $\eta > 0$. The system (1. 23) becomes in case $C = \mathcal{H}$ the "heavy ball method", sometimes called also "heavy ball method with friction". This nonlinear oscillator with damping is in case $\mathcal{H} = \mathbb{R}^2$ a simplified version of the differential system describing the motion of a heavy ball that rolls over the graph of g and that keep rolling under its own inertia until friction stop it at a critical point of g (see [17]).

The investigation of dynamical systems is motivated also by the fact that considering time discretization of these systems can lead to new discrete-type iterative schemes for solving monotone inclusion problems, a fact which has been underlined in the aforementioned papers. As an illustration, notice that the time discretization of (1. 21) leads to (1. 4) in case A is the (convex) subdifferential of Φ. For more on the relations between the continuous and discrete dynamics we refer the reader to [117].

Let us also mention that dynamical systems of implicit type have been considered in the literature also by Attouch and Svaiter in [20], Attouch, Abbas and Svaiter in [2] and Attouch, Alvarez and Svaiter in [18].

1.1 A description of the contents

In the following we give a description of the contents of this work, underlying its most important results. In Section 1.2 of the introduction we include several preliminary notions and results in order to make the manuscript as self-contained as possible.

Chapter 2. This chapter is dedicated to the investigation of the rate of convergence of a primal-dual splitting algorithm of forward-backward type introduced in [130] and designed to solve highly structured monotone inclusion problems as the ones described in (1. 10)-(1. 11).

In Section 2.1 we focus our attention on complexity results for the iterates generated by this algorithm. By incorporating variable step sizes, we succeed to accelerate the aforementioned algorithm and present two main results. For the first modified algorithm, by assuming that some of the operators involved are strongly monotone, we achieve for the sequence of primal iterates an order of convergence of $\mathcal{O}(\frac{1}{n})$. Further, under more involved strong monotonicity assumptions, we propose a second modified algorithm (this time with constant step sizes), which guarantees linear convergence for the sequence of both primal and dual iterates. We show how to particularize the general results in the context of nondifferentiable convex optimization problems (1. 12)-(1. 13), where some of the functions occurring in the objective are strongly convex. In the last part of Section 2.1 we present numerical experiments in image denoising and pattern recognition in cluster analysis and emphasize also the practical advantages of the modified iterative schemes over the initial one provided in [130]. Numerical comparisons to other state-of-the-art methods for convex nondifferentiable optimization problems are also made.

In Section 2.2 we investigate the rate of convergence of the sequence of objective function values of the algorithm given in [130] for the optimization problems (1. 12)-(1. 13). For the primal-dual splitting algorithms, mainly convergence statements for the sequence of iterates are available in the literature. However, especially from the point of view of solving real-life problems, the investigation of the convergence of the sequence of objective function values is of equal importance. We are able to prove a convergence rate of order $\mathcal{O}(\frac{1}{n})$ for the so-called primal-dual gap function attached to the pair of primal-dual problems. We illustrate this theoretical part by numerical experiments in image processing.

Chapter 3. In this chapter we carry out some investigations on inertial-type proximal-splitting algorithms.

In Section 3.1 we introduce and investigate the convergence properties of an inertial version of the Tseng's algorithm. We present first an inertial forward-backward-forward splitting algorithm for solving the monotone inclusion problem (1. 3) in case B is a (single-valued) monotone and Lipschitz continuous operator. The proposed scheme represents an extension of Tseng's forward-backward-forward-type method (see [62, 128, 129]) and for the study of its convergence properties we use some generalizations of the Fejér monotonicity techniques provided in [5]. Subsequently, we make use of the product space approach in order to obtain an inertial primal-dual splitting algorithm designed for solving monotone inclusion problems involving mixtures of linearly composed and parallel-sum type monotone operators, as considered in (1. 10)-(1. 11). We also show how the proposed iterative schemes can be used in order to solve primal-dual pairs of convex optimization problems of type (1. 12)-(1. 13).

In Section 3.2 we propose an inertial Douglas-Rachford proximal splitting algorithm. In order to prove its convergence we formulate first an 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 for monotone inclusions of type (1. 3) 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. Furthermore, we 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. We consider also the special case of primal-dual pairs of convex optimization problems and illustrate the theoretical results via some numerical experiments in clustering and location theory.

It is the aim of Section 3.3 to introduce and study the convergence properties of an inertial forward-backward proximal-type algorithm for the minimization of the sum of a nonsmooth and lower semicontinuous function and a smooth one in the full nonconvex setting. This scheme is characterized by the fact that, for the backward step we use a generalization of the proximal operator, not only by considering it to be, as it is natural in the nonconvex setting, a set-valued mapping, but also by replacing in its standard formulation the squared-norm by the Bregman distance of a strongly convex and differentiable function with Lipschitz continuous gradient. The techniques for proving the convergence of the numerical scheme use the same three main ingredients, as other proximal-type algorithms for nonconvex optimization problems given in the literature do. More precisely, we show a sufficient decrease property for the iterates, the existence of a subgradient lower bound for the iterates gap and, finally, we use the analytic features of the objective function in order to obtain convergence (see [11, 35]). The *limiting (Mordukhovich) subdifferential* and its properties play an important role in the analysis. The main result of this section shows that, provided an appropriate regularization of the objective satisfies the Kurdyka-Łojasiewicz property, the convergence of the inertial forward-backward algorithm is guaranteed. As a particular instance, we also treat the case when the objective function is semi-algebraic and present the convergence properties of the algorithm. In the last part of this section we consider two numerical experiments. The first one has an academic character and shows the ability of algorithms with inertial/memory effects to detect optimal solutions which are not found by their non-inertial versions (similar allegations can be found also in [110, Section 5.1] and [30, Example 1.3.9]). The second one concerns the restoration of a noisy blurred image by using a nonconvex misfit functional with nonconvex regularization.

Chapter 4. The aim of this chapter is to generalize the existing penalty-type splitting algorithms to the solving of more involved monotone inclusion problems.

In Section 4.1 we deal with problems of the form

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Dx + N_M(x), \tag{1. 24}$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ is a (single-valued) cocoercive operator and $M \subseteq \mathcal{H}$ is the (nonempty) set of zeros of another cocoercive operator $B : \mathcal{H} \to \mathcal{H}$. We propose a forward-backward penalty algorithm for solving (1. 24) and prove weak ergodic convergence for the generated sequence of iterates under a condition which involves the Fitzpatrick function associated to the operator B. Moreover, we prove strong convergence for the sequence of iterates whenever A is strongly monotone.

The investigations made are completed in Section 4.2 with the treatment of the monotone inclusion problem (1. 24), this time by relaxing the cocoercivity of D and B to monotonicity and Lipschitz continuity. We formulate in this more general setting a forward-backward-forward penalty type algorithm for solving (1. 24) and study its convergence properties. This study allows via some primal-dual techniques to deal with monotone inclusion problems having more complicated structures, for instance, involving mixtures of linearly composed maximally monotone operators and parallel-sum operators, like the one described in (1. 10)-(1. 11), but with an additional normal cone operator to the set of zeros of a single-valued mapping which is evaluated in the algorithm through a penalty term:

$$0 \in Ax + \sum_{i=1}^{m} L_i^*(B_i \Box D_i)(L_i x) + Cx + N_M(x). \tag{1. 25}$$

In the last part of the chapter we present these results in the context of solving convex minimization problems with intricate objective functions and consider a numerical example in image inpainting.

Chapter 5. We approach the solving of monotone inclusion problems of type (1. 3) in case B is single-valued by considering first and second order implicit-type dynamical systems.

We begin in Section 5.1 with the asymptotic analysis of a dynamical system associated with the fixed points set of a nonexpansive operator. While the existence of the trajectories of the ordinary differential equations is achieved in the framework of the Cauchy-Lipschitz-Picard Theorem, the (weak) convergence of the orbits to a fixed point of the operator is based on Lyapunov analysis combined with the continuous version of the Opial Lemma. We study also the convergence rates of the fixed point residual of the orbits of the dynamical system, for which we obtain a rate of convergence of order $o(1/\sqrt{t})$. Further, we propose also a generalization of the forward-backward continuous version of the dynamical system (1. 21) by considering instead of the convex subdifferential a maximally monotone operator A. This gives rise to the dynamical system

$$\left\{ \begin{array}{l} \dot{x}(t) = \lambda(t) \left[J_{\eta A}\Big(x(t) - \eta B(x(t)) \Big) - x(t) \right] \\ x(0) = x_0, \end{array} \right. \tag{1. 26}$$

which we associate with the inclusion problem (1. 3). In the last part of this section we show that the trajectory of (1. 26) strongly converges with exponential rate to the unique solution of (1. 3), provided the sum $A + B$ is strongly monotone.

In Section 5.2 we investigate second order dynamical systems associated to monotone inclusion problems. We start with ordinary differential equations associated to the set of zeros of a cocoercive operator. We distinguish between anisotropic damping parameters induced by an elliptic operator as in [3] and time depended damping parameters. Further, we approach the problem (1. 3) by the dynamical system

$$\left\{ \begin{array}{l} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t) \left[x(t) - J_{\eta A}\Big(x(t) - \eta B(x(t)) \Big) \right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{array} \right. \tag{1. 27}$$

We specialize these investigations to the minimization of the sum of a nonsmooth convex function with a smooth convex function one, fact which allows us to recover and improve results given in [7, 8] in the context of studying the dynamical system (1. 23). When considering the unconstrained minimization of a smooth convex function we prove a rate of $\mathcal{O}(1/t)$ for the convergence of the function value along the ergodic trajectory to its minimum value. The last part of this section is dedicated to convergence rates for strongly monotone inclusions. By weakening the assumptions on B to monotonicity and Lipschitz continuity, however, provided that $A+B$ is strongly monotone, the trajectories of (1. 27) converge strongly to the unique zero of $A+B$ with an exponential rate. Exponential convergence rates have been obtained also by Antipin in [7] for the dynamical systems (1. 19) and (1. 23), by imposing for the smooth function g supplementary strong convexity assumptions. We derive from here convergence rates for the trajectories generated by dynamical systems associated to the minimization of the sum of a proper, convex and lower semicontinuous function with a smooth convex one provided the objective function fulfills a strong convexity assumption. In the particular case of minimizing a smooth and strongly convex function, we prove that its values converge along the trajectory to its minimum value with exponential rate, too.

1.2 Preliminary notions and results

This section is dedicated to the presentation of several notations and results which are used throughout the manuscript. Some (technical) results or notions which are specific only to some sections are presented where needed. We refer the reader to [26, 37, 38, 85, 124, 131] for standard notations in monotone operator theory and convex analysis.

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 weak and strong convergence, respectively. The following identity will be used several times (see for example [26, Corollary 2.14]):

$$\|\alpha x+(1-\alpha)y\|^2+\alpha(1-\alpha)\|x-y\|^2 = \alpha\|x\|^2+(1-\alpha)\|y\|^2 \ \forall\alpha\in\mathbb{R}\ \forall(x,y)\in\mathcal{H}\times\mathcal{H}. \tag{1. 28}$$

When $\mathcal{G}$ is another Hilbert space and $K:\mathcal{H}\rightarrow\mathcal{G}$ a linear continuous operator, then the *norm* of K is defined as $\|K\| = \sup\{\|Kx\| : x\in\mathcal{H}, \|x\|\leq 1\}$, while $K^*:\mathcal{G}\rightarrow\mathcal{H}$, 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 $S\subseteq\mathcal{H}$ a convex set, we denote by

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

its *strong quasi-relative interior*. Notice that we always have $\operatorname{int} S\subseteq\operatorname{sqri} S$ (in general this inclusion may be strict). If $\mathcal{H}$ is finite-dimensional, then $\operatorname{sqri} S$ coincides with $\operatorname{ri} S$, the relative interior of S, which is the interior of S with respect to its affine hull. The notion of strong quasi-relative interior belongs to the class of generalized interiority notions and play an important role in the formulation of regularity conditions which are used in the theory of convex optimization problems in order to guarantee *strong duality*, namely the situation when the optimal objective values of the primal optimization problem and its dual one coincide and the dual has an optimal solution. We refer to [26, 37, 38, 40, 85, 124, 131] for other interiority notions and their impact in the duality theory.

An efficient tool for proving weak convergence of a sequence in Hilbert spaces (without a priori knowledge of the limit) is the celebrated Opial Lemma, which we recall in the following.

Lemma 1.1 *(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 weak sequential 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.

In order to prove the first part of the Opial Lemma, one usually tries to show that the sequence $(\|x_n - x\|)_{n\in\mathbb{N}}$, where $x \in C$, fulfills a Fejér-type inequality. In this sense the following result is very useful.

Lemma 1.2 *Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ be real sequences. Assume that $(a_n)_{n\in\mathbb{N}}$ is bounded from below, $(b_n)_{n\in\mathbb{N}}$ is nonnegative, $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1$ and $a_{n+1} - a_n + b_n \leq \varepsilon_n$ for all $n \in \mathbb{N}$. Then $(a_n)_{n\in\mathbb{N}}$ is convergent and $(b_n)_{n\in\mathbb{N}} \in \ell^1$.*

Let us recall now some facts about monotone operators. 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*
- $\operatorname{dom} A = \{x \in \mathcal{H} : Ax \neq \emptyset\}$ its *domain*
- $\operatorname{ran} A = \cup_{x\in\mathcal{H}} Ax$ its *range*
- $A^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ its *inverse operator*, defined by $(u, x) \in \operatorname{gr} A^{-1}$ if and only if $(x, u) \in \operatorname{gr} A$
- $\operatorname{zer} A = \{x \in \mathcal{H} : 0 \in Ax\}$ the *set of zeros* of the operator A.

We say that A is *monotone* if

$$\langle x - y, u - v\rangle \geq 0 \ \forall (x, u), (y, v) \in \operatorname{gr} A.$$

A monotone operator A is said to be *maximally monotone*, if there exists no proper monotone extension of the graph of A on $\mathcal{H} \times \mathcal{H}$. Let us mention that in case A is maximally monotone, $\operatorname{zer} A$ is a convex and closed set [26, Proposition 23.39]. We refer to [26, Section 23.4] for conditions ensuring that $\operatorname{zer} A$ is nonempty. If A is maximally monotone, then one has the following characterization for the set of its zeros:

$$z \in \operatorname{zer} A \text{ if and only if } \langle u - z, w\rangle \geq 0 \text{ for all } (u, w) \in \operatorname{gr} A. \tag{1. 29}$$

The *resolvent* of A, $J_A : \mathcal{H} \rightrightarrows \mathcal{H}$, is defined by

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

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 (asee [26, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [26, 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 [26, Proposition 23.18])

$$J_{\gamma A} + \gamma J_{\gamma^{-1} A^{-1}} \circ \gamma^{-1} \operatorname{Id}_{\mathcal{H}} = \operatorname{Id}_{\mathcal{H}}. \tag{1. 30}$$

The resolvent operator will play an important role in the formulation of the algorithms and dynamical systems considered in connection with determining the set of zeros of (sums) of monotone operators.

Further, let us mention some classes of operators that are used in the following. We say that A is *demiregular* at $x \in \operatorname{dom} A$, if for every sequence $(x_n, u_n)_{n\in\mathbb{N}} \in \operatorname{gr} A$ and every $u \in Ax$ such that $x_n \rightharpoonup x$ and $u_n \to u$, we have $x_n \to x$. We refer the reader to [12, Proposition 2.4] and [62, Lemma 2.4] for conditions ensuring this property. The operator A is said to be *uniformly monotone* at $x \in \operatorname{dom} A$, if there exists an increasing function $\phi_A : [0, +\infty) \to [0, +\infty]$ that vanishes only at 0, and

$$\langle x - y, u - v\rangle \geq \phi_A(\|x - y\|) \ \ \forall u \in Ax \text{ and } \forall (y, v) \in \operatorname{gr} A.$$

If this inequality holds for all $(x, u), (y, v) \in \operatorname{gr} A$, we say that A is uniformly monotone. Let us mention that, if A is uniformly monotone at $x \in \operatorname{dom} A$, then it is demiregular at x.

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

$$\langle x - y, u - v\rangle \geq \gamma\|x - y\|^2 \ \forall (x, u), (y, v) \in \operatorname{gr} A. \tag{1. 31}$$

Notice that if A is maximally monotone and strongly monotone, then $\operatorname{zer} A$ is a singleton, thus nonempty (see [26, Corollary 23.37]). A single-valued operator $A : \mathcal{H} \to \mathcal{H}$ is said to be γ-*cocoercive*, if

$$\langle x - y, Ax - Ay\rangle \geq \gamma\|Ax - Ay\|^2 \ \forall (x, y) \in \mathcal{H} \times \mathcal{H}.$$

Moreover, A is γ-*Lipschitz continuous*, if $\|Ax-Ay\| \leq \gamma\|x-y\|$ for all $(x, y) \in \mathcal{H}\times\mathcal{H}$. A single-valued linear operator $A : \mathcal{H} \to \mathcal{H}$ is said to be *skew*, if $\langle x, Ax\rangle = 0$ for all $x \in \mathcal{H}$.

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 [26, Corollary 23.10]).

The following result, which is a consequence of the demiclosedness principle (see [26, Theorem 4.17]), will be useful in the proof of the convergence of the inertial version of the Krasnosel'skiĭ–Mann algorithm in Chapter 3. It will be used also in Chapter 5 in the context of studying dynamical systems associated with the fixed point set of a nonexpansive operator.

Lemma 1.3 *(see [26, 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 $x \in \operatorname{Fix} T$.*

We recall also the following subclass of the nonexpansive operators. Let $\alpha \in (0, 1)$ be fixed. We say that $R : \mathcal{H} \to \mathcal{H}$ is α-*averaged*, if there exists a nonexpansive operator $T : \mathcal{H} \to \mathcal{H}$ such that $R = (1 - \alpha)\operatorname{Id} + \alpha T$. For $\alpha = \frac{1}{2}$ we obtain as an important representative of this class the *firmly nonexpansive* operators. For properties and other insides concerning these families of operators we refer to [26].

Finally, the *parallel sum* of two operators $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined by $A \Box B : \mathcal{H} \rightrightarrows \mathcal{H}$

$$A \Box B = (A^{-1} + B^{-1})^{-1}.$$

In the last part of this section we recall some elements of convex analysis. For a function $f : \mathcal{H} \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ is the extended real line, we denote by

$$\operatorname{dom} f = \{x \in \mathcal{H} : f(x) < +\infty\}$$

its *effective domain* and say that f is *proper* if $\operatorname{dom} f \neq \emptyset$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. Concerning calculus rules where $\pm\infty$ are involved we make the following conventions (see [131]): $(+\infty) + (-\infty) = +\infty$, $0(+\infty) = +\infty$ and $0(-\infty) = 0$. We denote by $\Gamma(\mathcal{H})$ the family of proper, convex and lower semi-continuous extended real-valued functions defined on $\mathcal{H}$. Let $f^* : \mathcal{H} \to \overline{\mathbb{R}}$,

$$f^*(u) = \sup_{x \in \mathcal{H}} \{\langle u, x\rangle - f(x)\} \ \forall u \in \mathcal{H},$$

denote the *conjugate function* of f. We also denote by $\min f := \inf_{x \in \mathcal{H}} f(x)$ and by $\operatorname{argmin} f := \{x \in \mathcal{H} : f(x) = \min f\}$.

The (convex) *subdifferential* of f is a set-valued operator $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ defined by

$$\partial f(x) = \begin{cases} \{v \in \mathcal{H} : f(y) \geq f(x) + \langle v, y - x\rangle \ \forall y \in \mathcal{H}\}, & \text{if } f(x) \in \mathbb{R}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Let us mention that if f is proper, convex, and Fréchet differentiable at $\overline{x}$, then $\partial f(\overline{x}) = \{\nabla f(\overline{x})\}$ (cf. [131, Corollary 2.4.10 and Theorem 2.4.4(i)]). The Fermat rule in the nondifferentiable case underlines the usefulness of the subdifferential: if f is proper, then for $\overline{x} \in \operatorname{dom} f$ we have the relation

$$\overline{x} \in \operatorname{argmin} f \Leftrightarrow 0 \in \partial f(\overline{x}).$$

Notice that if $f \in \Gamma(\mathcal{H})$, then ∂f is a maximally monotone operator (cf. [120]) and it holds $(\partial f)^{-1} = \partial f^*$. For $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ two proper functions, we consider their *infimal convolution*, which is the function $f \Box g : \mathcal{H} \to \overline{\mathbb{R}}$, defined by

$$(f \Box g)(x) = \inf_{y \in \mathcal{H}} \{f(y) + g(x - y)\} \ \forall x \in \mathcal{H}.$$

In case $f, g \in \Gamma(\mathcal{H})$ and a regularity condition is fulfilled, according to [26, Proposition 24.27] we have $\partial f \Box \partial g = \partial(f \Box g)$, and this justifies the notation used for the parallel sum of two operators as described above.

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

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

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

Notice that we have the following formula for the resolvent of the subdifferential operator:

$$J_{\gamma \partial f} = (\operatorname{Id}_{\mathcal{H}} + \gamma \partial f)^{-1} = \operatorname{prox}_{\gamma f},$$

thus $\operatorname{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H}$ is a single-valued operator fulfilling the extended *Moreau's decomposition formula*

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

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 $\phi = (\gamma/2)(\cdot)^2$, where $\gamma > 0$, then f is said to be *γ-strongly convex*. Let us mention that this property implies γ-strong monotonicity of ∂f (see [26, Example 22.3]) (more general, if f is uniformly convex, then ∂f is uniformly monotone, see [26, Example 22.3]). Furthermore, the proper function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is γ-strongly convex, if and only if $f - \frac{\gamma}{2}\| \cdot \|^2$ is a convex function. We mention also the following interesting connection between the strong convexity of a function and the differentiability properties of its conjugate: if $f \in \Gamma(\mathcal{H})$, then f is γ-strongly convex if and only if f^* is Fréchet differentiable with γ^{-1}-Lipschitz continuous gradient (see [26, Theorem 18.15], [131, Corollary 3.5.11, Remark 3.5.3]).

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

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

where $\operatorname{proj}_S : \mathcal{H} \to S$ denotes the *projection operator* on S (see [26, Example 23.3 and Example 23.4]).

Finally, the descent lemma which we recall below will be used several times as a technical tool in order to derive useful inequalities in the convergence analysis of the algorithms and dynamical systems proposed in the manuscript.

Lemma 1.4 *(see [106, Lemma 1.2.3]) Let $g : \mathcal{H} \to \mathbb{R}$ be Fréchet differentiable with L-Lipschitz continuous gradient. Then*

$$g(y) \leq g(x) + \langle \nabla g(x), y - x \rangle + \frac{L}{2}\|y - x\|^2 \ \forall x, y \in \mathcal{H}.$$

Chapter 2

Complexity results for a primal-dual splitting algorithm of forward-backward type

The main goal of this chapter is to present several convergence rates related to a primal-dual splitting algorithm of forward-backward-type associated with highly structured monotone inclusion problems. While in Section 2.1 we focus our attention on complexity results concerning the iterates in case of strongly monotone inclusion problems, in Section 2.2 we consider the case of convex optimization problems with intricate objective functions and give a rate of convergence for the objective function values.

We are concerned with the study of the convergence rate of a primal-dual splitting algorithm introduced in [130]. The following problem represents the starting point of our investigations.

Problem 2.1 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ an η-cocoercive operator for $\eta > 0$. Let m be a strictly positive integer and, for every $i \in \{1,...,m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone operator, let $D_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone and ν_i-strongly monotone operator for $\nu_i > 0$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion*

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

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \left\{ \begin{array}{l} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in (B_i \square D_i)(L_i x - r_i),\ i = 1,...,m. \end{array} \right. \tag{2. 2}$$

We say that $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to Problem 2.1, if

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in A\overline{x} + C\overline{x} \text{ and } \overline{v}_i \in (B_i \square D_i)(L_i\overline{x} - r_i),\ i = 1,...,m. \tag{2. 3}$$

If $\overline{x} \in \mathcal{H}$ is a solution to (2. 1), 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 2.1 and, if $(\overline{v}_1,...,\overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a solution to (2. 2), 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 2.1. Moreover, if $(\overline{x}, \overline{v}_1,...,\overline{v}_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to Problem 2.1, then $\overline{x}$ is a solution to (2. 1) and $(\overline{v}_1,...,\overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a solution to (2. 2).

By employing the classical forward-backward algorithm (see for example [75, 129]) in a renormed product space, Vũ proposed in [130] an iterative scheme for solving a slightly modified version of Problem 2.1 formulated in the presence of some given weights $w_i \in (0, 1]$, $i = 1,..., m$, with $\sum_{i=1}^m w_i = 1$ for the terms occurring in the second summand of the primal inclusion problem. The following result is an adaption of [130, Theorem 3.1] to Problem 2.1 in the error-free case and when $\lambda_n = 1$ for all $n \geq 0$. Let us mention that under a different approach which relies on Fejér monotonicity techniques, the convergence of an equivalent form of the algorithm presented in the theorem below has been investigated in [52] for monotone inclusions of less involved structures as the ones considered in (2. 5)-(2. 6), where one additionally assumes that $Cx = 0$ for all $x \in \mathcal{H}$.

Theorem 2.1 *(see [130]) In Problem 2.1 suppose that*

$$z \in \operatorname{ran}\left(A + \sum_{i=1}^m L_i^*\big((B_i \Box D_i)(L_i \cdot - r_i)\big) + C\right). \tag{2. 4}$$

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

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

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

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

Then there exists a primal-dual solution $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ to Problem 2.1 such that the following statements are true:

(a) $x_n \rightharpoonup \overline{x}$ *and* $(v_{1,n},..., v_{m,n}) \rightharpoonup (\overline{v}_1,..., \overline{v}_m)$ *as* $n \to +\infty$;

(b) if C is uniformly monotone, then $x_n \to \overline{x}$, as $n \to +\infty$;

(c) if D_i^{-1} is uniformly monotone for some $i \in \{1,..., m\}$, then $v_{i,n} \to \overline{v}_i$ as $n \to +\infty$.

Remark 2.1 Notice that the work in [130] is closely related to [69] and [79], where primal-dual splitting methods for nonsmooth convex optimization problems are proposed. More exactly, the convergence property of [69, Algorithm 1] proved in [69, Theorem 1] follow as special instance of the main result in [130]. On the other hand, Condat proposes in [79] an algorithm which can be seen as an extension of the one in [69] to optimization problems in the objective of which convex differentiable functions occur, as well.

Remark 2.2 We would like to stress the fact that the relation (2. 4) is equivalent to the existence of primal-dual solutions to Problem 2.1 above.

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

In this section we propose under appropriate strong monotonicity assumptions two modified versions of the algorithm in Theorem 2.1. The first one ensures an order of convergence of $\mathcal{O}(\frac{1}{n})$ for the sequences of primal iterates, while the second one, under more involved strong monotonicity conditions, guarantees linear convergence for the sequences of primal and dual iterates.

2.1.1 The case $A + C$ is strongly monotone

For the beginning, we show that, in case $A + C$ is strongly monotone, one can guarantee an order of convergence of $\mathcal{O}(\frac{1}{n})$ for the sequence $(x_n)_{n\geq 0}$. To this end, we update in each iteration the parameters τ and σ_i, $i = 1,..., m$, and use a modified formula for the sequence $(y_n)_{n\geq 0}$. Let us notice that incorporating variable step sizes can also increase the dynamic of the sequences involved, with possible numerical performances, as underlined also in [132] and [69].

Due to technical reasons, we apply this method in case D_i^{-1} is equal to zero for $i = 1,..., m$, that is $D_i(0) = \mathcal{G}_i$ and $D_i(x) = \emptyset$ for $x \neq 0$. Let us notice that, by using the approach proposed in [58, Remark 3.2], one can extend the statement of Theorem 2.2 below, which is the main result of this subsection, to the primal-dual pair of monotone inclusions as stated in Problem 2.1.

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

Problem 2.2 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ a monotone and η-Lipschitz continuous operator for $\eta > 0$. Let m be a strictly positive integer and, for every $i \in \{1,..., m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone operator and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion*

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^*(B_i(L_i\overline{x} - r_i)) + C\overline{x}, \tag{2. 5}$$

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \begin{cases} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in B_i(L_ix - r_i),\ i = 1,..., m. \end{cases} \tag{2. 6}$$

As for Problem 2.1, we say that $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to Problem 2.2, if

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in A\overline{x} + C\overline{x} \text{ and } \overline{v}_i \in B_i(L_i\overline{x} - r_i),\ i = 1,..., m. \tag{2. 7}$$

Remark 2.3 One can notice that, in comparison to Problem 2.1, we relax in Problem 2.2 the assumptions made on the operator C. It is obvious that, if C is a η-cocoercive operator for $\eta > 0$, then C is monotone and $1/\eta$-Lipschitz continuous. Although in case C is the gradient of a convex and differentiable function, due to the celebrated Baillon-Haddad Theorem (see, for instance, [26, Corollary 8.16]), the two classes of operators coincide, the second one is in general larger. Indeed, nonzero linear, skew and Lipschitz continuous operators are not cocoercive. For example, when $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces and $L : \mathcal{H} \to \mathcal{G}$ is nonzero linear continuous, then $(x, v) \mapsto (L^*v, -Lx)$ is an operator having all these properties. This operator appears in a natural way when employing primal-dual approaches in the context of monotone inclusion problems as done in [62] (see also [52, 58, 76, 130]).

Under the assumption that $A + C$ is γ-strongly monotone for $\gamma > 0$ we propose the following modification of the iterative scheme in Theorem 2.1.

Algorithm 2.1

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

Remark 2.4 Notice that in contrast to the algorithm in Theorem 2.1, we allow here variable step sizes τ_n and $\sigma_{i,n}$, $1 = 1, ..., m$, which are updated in each iteration. Moreover, for every $n \geq 0$, the iterate y_n is defined by means of the sequence θ_n. Dynamic step sizes have been first proposed in [132] and then used in [69] in order to accelerate the convergence of iterative methods when solving convex optimization problems.

Remark 2.5 The assumption $\tau_0 \sum_{i=1}^m \sigma_{i,0} \|L_i\|^2 \leq \sqrt{1 + \tau_0(2\gamma - \eta\tau_0)/\lambda}$ in Algorithm 2.1 is equivalent to $\tau_1 \sum_{i=1}^m \sigma_{i,0} \|L_i\|^2 \leq 1$, being fulfilled if $\tau_0 > 0$ is chosen such that

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

We present now the following complexity result of the algorithm described above.

Theorem 2.2 *Suppose that $A + C$ is γ-strongly monotone for $\gamma > 0$ and let $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ be a primal-dual solution to Problem 2.2. Then the sequences generated by Algorithm 2.1 fulfill for all $n \geq 0$*

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

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

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

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

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

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

Since $A+C$ is γ-strongly monotone, (2. 7) and (2. 9) yield for all $n \geq 0$

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

Further, we have

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

and, since C is η-Lipschitz continuous,

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

Hence, it follows from (2. 10)–(2. 11) and the last inequality that for all $n \geq 0$ it holds

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

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

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

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

$$\frac{1}{\sigma_{i,n}}(v_{i,n} - v_{i,n+1}) + L_i y_n - r_i \in B_i^{-1} v_{i,n+1}, \quad (2.\ 13)$$

the monotonicity of B_i^{-1} and (2. 7), we obtain

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

hence

$$\frac{\|v_{i,n+1} - \overline{v}_i\|^2}{\sigma_{i,n}} \leq \frac{\|v_{i,n} - \overline{v}_i\|^2}{\sigma_{i,n}} - \frac{\|v_{i,n} - v_{i,n+1}\|^2}{\sigma_{i,n}} + 2\langle L_i(y_n - \overline{x}), v_{i,n+1} - \overline{v}_i\rangle. \quad (2.\ 14)$$

Summing up the inequalities in (2. 12) and (2. 14) we obtain for all $n \geq 0$

$$\begin{aligned} &\left(\frac{\lambda}{\tau_{n+1}} + 2\gamma - \eta\tau_{n+1}\right)\|x_{n+2} - \overline{x}\|^2 + \sum_{i=1}^{m}\frac{\|v_{i,n+1} - \overline{v}_i\|^2}{\sigma_{i,n}} \\ &\leq \frac{\lambda}{\tau_{n+1}}\|x_{n+1} - \overline{x}\|^2 + \sum_{i=1}^{m}\frac{\|v_{i,n} - \overline{v}_i\|^2}{\sigma_{i,n}} \\ &\quad - \frac{\|x_{n+2} - x_{n+1}\|^2}{\tau_{n+1}} - \sum_{i=1}^{m}\frac{\|v_{i,n} - v_{i,n+1}\|^2}{\sigma_{i,n}} \\ &\quad + 2\sum_{i=1}^{m}\langle L_i(x_{n+2} - y_n), -v_{i,n+1} + \overline{v}_i\rangle. \end{aligned} \quad (2.\ 15)$$

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

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

By combining the last inequality with (2. 15), we obtain for all $n \geq 0$

$$\begin{aligned}
&\left(\frac{\lambda}{\tau_{n+1}} + 2\gamma - \eta\tau_{n+1}\right)\|x_{n+2} - \overline{x}\|^2 + \sum_{i=1}^{m} \frac{\|v_{i,n+1} - \overline{v}_i\|^2}{\sigma_{i,n}} + \frac{\|x_{n+2} - x_{n+1}\|^2}{\tau_{n+1}} \\
&-2\sum_{i=1}^{m}\langle L_i(x_{n+2} - x_{n+1}), -v_{i,n+1} + \overline{v}_i\rangle \qquad (2.\ 16)\\
&\leq \frac{\lambda}{\tau_{n+1}}\|x_{n+1} - \overline{x}\|^2 + \sum_{i=1}^{m} \frac{\|v_{i,n} - \overline{v}_i\|^2}{\sigma_{i,n}} + \left(\sum_{i=1}^{m}\|L_i\|^2\sigma_{i,n}\right)\theta_n^2\|x_{n+1} - x_n\|^2 \\
&\quad -2\sum_{i=1}^{m}\theta_n\langle L_i(x_{n+1} - x_n), -v_{i,n} + \overline{v}_i\rangle.
\end{aligned}$$

After dividing (2. 16) by τ_{n+1} and noticing that for all $n \geq 0$

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

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

and

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

it follows that the Fejér-type inequality (2. 8) is true.

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

$$\begin{aligned}
&\frac{\lambda}{\tau_{N+1}^2}\|x_{N+1} - \overline{x}\|^2 + \sum_{i=1}^{m} \frac{\|v_{i,N} - \overline{v}_i\|^2}{\tau_1\sigma_{i,0}} + \frac{\|x_{N+1} - x_N\|^2}{\tau_N^2} \\
&\leq \frac{\lambda}{\tau_1^2}\|x_1 - \overline{x}\|^2 + \sum_{i=1}^{m} \frac{\|v_{i,0} - \overline{v}_i\|^2}{\tau_1\sigma_{i,0}} + \frac{\|x_1 - x_0\|^2}{\tau_0^2} \qquad (2.\ 17)\\
&\quad +2\sum_{i=1}^{m}\left(\frac{1}{\tau_N}\langle L_i(x_{N+1} - x_N), -v_{i,N} + \overline{v}_i\rangle - \frac{1}{\tau_0}\langle L_i(x_1 - x_0), -v_{i,0} + \overline{v}_i\rangle\right).
\end{aligned}$$

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

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

and obtain from (2. 17) that

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

which rapidly yields 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 all $n \geq 0$

$$\tau_{n+1} = \frac{\tau_n}{\sqrt{1 + \frac{\tau_n}{\lambda}(2\gamma - \eta\tau_n)}}. \tag{2. 18}$$

Since $0 < \tau_0 < 2\gamma/\eta$, it follows by induction that $0 < \tau_{n+1} < \tau_n < \tau_0 < 2\gamma/\eta$ for all $n \geq 1$, hence the sequence $(\tau_n)_{n\geq 0}$ converges. In the light of (2. 18) 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

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

$\square$

Remark 2.6 Let us mention that, if $A + C$ is γ-strongly monotone with $\gamma > 0$, then the operator $A + \sum_{i=1}^m L_i^*(B_i(L_i \cdot -r_i)) + C$ is strongly monotone, as well, thus the monotone inclusion problem (2. 5) has at most one solution. Hence, if $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ is a primal-dual solution to Problem 2.2, then $\overline{x}$ is the unique solution to (2. 5). Notice that the problem (2. 6) may have more than one solution.

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

In this subsection we propose a further modified version of the algorithm in Theorem 2.1. The main result of this section is that if $A + C$ and $B_i^{-1} + D_i^{-1}$, $i = 1,..., m$, are strongly monotone, then one achieves linear convergence rate for the sequences $(x_n)_{n\geq 0}$ and $(v_{i,n})_{n\geq 0}$, $i = 1,..., m$. The algorithm aims to solve the primal-dual pair of monotone inclusions stated in Problem 2.1 under relaxed assumptions for the operators C and D_i^{-1}, $i = 1,..., m$. A same comment as in Remark 2.9 can be made also in this context.

Problem 2.3 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ a monotone and η-Lipschitz continuous operator for $\eta > 0$. Let m be a strictly positive integer and, for every $i \in \{1,..., m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone operator, let $D_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a monotone operator such that D_i^{-1} is ν_i-Lipschitz continuous for $\nu_i > 0$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion*

$$\textit{find } \overline{x} \in \mathcal{H} \textit{ such that } z \in A\overline{x} + \sum_{i=1}^m L_i^*\big((B_i \Box D_i)(L_i\overline{x} - r_i)\big) + C\overline{x}, \tag{2. 19}$$

together with the dual inclusion

$$\textit{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \textit{ such that } \exists x \in \mathcal{H} : \left\{ \begin{array}{l} z - \sum_{i=1}^m L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in (B_i \Box D_i)(L_i x - r_i),\ i = 1,..., m. \end{array} \right. \tag{2. 20}$$

Under the assumption that $A+C$ is γ-strongly monotone for $\gamma>0$ and $B_i^{-1}+D_i^{-1}$ is δ_i-strongly monotone with $\delta_i>0$, $i=1,....m$, we propose the following modification of the iterative scheme in Theorem 2.1.

Algorithm 2.2

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

Remark 2.7 Different to Algorithm 2.1, the step sizes are now constant in each iteration, as it is also the case in Theorem 2.1. The major difference to the iterative scheme in Theorem 2.1 is given by the role played by the constant μ, not only in the definition of the step sizes, but also in the way the sequence $(y_n)_{n\geq 0}$ is constructed (through the choice of θ). Notice that the situation when $\theta=1$ provides the same definition of the latter as in the algorithm stated in Theorem 2.1.

Theorem 2.3 *Suppose that $A+C$ is γ-strongly monotone for $\gamma>0$, $B_i^{-1}+D_i^{-1}$ is δ_i-strongly monotone for $\delta_i>0$, $i=1,...,m$, and let $(\overline{x},\overline{v}_1,...,\overline{v}_m)$ be a primal-dual solution to Problem 2.3 (that is (2. 3) holds). Then the sequences generated by Algorithm 2.2 fulfill for all $n\geq 0$*

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

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

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

$$\frac{1}{\tau}(x_{n+1}-x_{n+2})-\left(\sum_{i=1}^m L_i^*v_{i,n+1}+Cx_{n+1}-z\right)+Cx_{n+2}\in(A+C)x_{n+2}, \tag{2. 21}$$

thus, since $A+C$ is γ-strongly monotone, (2. 20) yields

$$\begin{aligned}&\gamma\|x_{n+2}-\overline{x}\|^2\leq\left\langle x_{n+2}-\overline{x},\frac{1}{\tau}(x_{n+1}-x_{n+2})\right\rangle\\&+\left\langle x_{n+2}-\overline{x},-\left(\sum_{i=1}^m L_i^*v_{i,n+1}+Cx_{n+1}-z\right)+Cx_{n+2}-\left(z-\sum_{i=1}^m L_i^*\overline{v}_i\right)\right\rangle\\&=\frac{1}{\tau}\langle x_{n+2}-\overline{x},x_{n+1}-x_{n+2}\rangle+\langle x_{n+2}-\overline{x},Cx_{n+2}-Cx_{n+1}\rangle\\&\quad+\sum_{i=1}^m\langle L_i(x_{n+2}-\overline{x}),\overline{v}_i-v_{i,n+1}\rangle\,.\end{aligned}\tag{2. 22}$$

Further, by using (2. 11) and

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

(which is a consequence of the Lipschitz property of the operator C), we get from (2. 22) that for all $n \geq 0$

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

After multiplying this inequality with μ and taking into account that

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

we obtain for any $n \geq 0$

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

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

$$\frac{1}{\sigma_i}(v_{i,n}-v_{i,n+1})+L_iy_n-D_i^{-1}v_{i,n}-r_i+D_i^{-1}v_{i,n+1}\in(B_i^{-1}+D_i^{-1})v_{i,n+1}, \quad (2.\ 24)$$

the δ_i-strong monotonicity of $B_i^{-1}+D_i^{-1}$ and (2. 20), we obtain

$$\delta_i\|v_{i,n+1}-\overline{v}_i\|^2\leq\left\langle\frac{1}{\sigma_i}(v_{i,n}-v_{i,n+1}),v_{i,n+1}-\overline{v}_i\right\rangle + \left\langle L_iy_n-r_i-D_i^{-1}v_{i,n}+D_i^{-1}v_{i,n+1}-(L_i\overline{x}-r_i),v_{i,n+1}-\overline{v}_i\right\rangle. \quad (2.\ 25)$$

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

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

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

$$\langle D_i^{-1}v_{i,n+1}-D_i^{-1}v_{i,n},v_{i,n+1}-\overline{v}_i\rangle\leq\frac{\delta_i}{2}\|v_{i,n+1}-\overline{v}_i\|^2+\frac{\nu_i^2}{2\delta_i}\|v_{i,n+1}-v_{i,n}\|^2. \quad (2.\ 26)$$

Consequently, from (2. 25) and(2. 26) we obtain for every $i = 1,...,m$ and every $n \geq 0$:

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

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

$$\delta_i\left(1+\frac{\mu}{2}\right)\|v_{i,n+1}-\overline{v}_i\|^2\leq\delta_i\|v_{i,n}-\overline{v}_i\|^2-\frac{\delta_i}{2}\|v_{i,n+1}-v_{i,n}\|^2 +\mu\langle L_i(\overline{x}-y_n),\overline{v}_i-v_{i,n+1}\rangle. \quad (2.\ 27)$$

We denote

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

Summing up the inequalities in (2. 23) and (2. 27), we obtain for all $n \geq 0$

$$\begin{aligned}\left(1+\frac{\mu}{2}\right)a_{n+1} \leq a_n - \frac{\gamma}{2}\|x_{n+2}-x_{n+1}\|^2 - \sum_{i=1}^m \frac{\delta_i}{2}\|v_{i,n}-v_{i,n+1}\|^2 \\ + \mu\sum_{i=1}^m \langle L_i(x_{n+2}-y_n), \overline{v}_i - v_{i,n+1}\rangle. \qquad (2.\ 28)\end{aligned}$$

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

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

Taking into consideration that

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

from (2. 28), we obtain for all $n \geq 0$

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

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

$$\begin{aligned}&\omega^{-(n+1)}a_{n+1} + \frac{\gamma}{2}\omega^{-n}\|x_{n+2}-x_{n+1}\|^2 \\ &+ \mu\omega^{-n}\sum_{i=1}^m \langle L_i(x_{n+2}-x_{n+1}), v_{i,n+1}-\overline{v}_i\rangle \\ &\leq \omega^{-n}a_n + \frac{\gamma}{2}\omega^{-(n-1)}\|x_{n+1}-x_n\|^2 \\ &\quad + \mu\omega^{-(n-1)}\sum_{i=1}^m \langle L_i(x_{n+1}-x_n), v_{i,n}-\overline{v}_i\rangle. \qquad (2.\ 29)\end{aligned}$$

Let $N \in \mathbb{N}, N \geq 2$. Summing up the inequality in (2. 29) from $n = 0$ to $N-1$, it yields

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

Using that

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

this further yields

$$\omega^{-N}a_N + \omega^{-N+1}\left(\frac{\gamma}{2} - \frac{\mu^2}{4}\sum_{i=1}^{m}\frac{\|L_i\|^2}{\delta_i}\right)\|x_N - x_{N+1}\|^2$$
$$-\omega^{-N+1}\sum_{i=1}^{m}\delta_i\|v_{i,N}-\overline{v}_i\|^2$$
$$\leq a_0 + \frac{\gamma}{2}\omega\|x_1 - x_0\|^2 + \mu\omega\sum_{i=1}^{m}\langle L_i(x_1 - x_0), v_{i,0} - \overline{v}_i\rangle. \tag{2. 30}$$

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

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

hence, after multiplying (2. 30) with ω^{-N}, it yields

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

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

Remark 2.8 If $A + C$ is γ-strongly monotone for $\gamma > 0$ and $B_i^{-1} + D_i^{-1}$ is δ_i-strongly monotone for $\delta_i > 0$, $i = 1,..., m$, then there exists at most one primal-dual solution to Problem 2.3. Hence, if $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ is a primal-dual solution to Problem 2.3, then $\overline{x}$ is the unique solution to the primal inclusion (2. 19) and $(\overline{v}_1,..., \overline{v}_m)$ is the unique solution to the dual inclusion (2. 20).

Remark 2.9 The modified versions Algortihm 2.1 and Algorithm 2.2 can handle Problem 2.1 under more general hypotheses than the original method given in [130]. Indeed, convergence was shown under more general hypotheses on the operator C for the first (see also Remark 2.3) and on the operators $D_i, i = 1, ..., m$ for the latter. More than that, we can provide in both cases a rate of convergence for the sequence of the primal iterates and in case of Algorithm 2.2 one for the sequence of dual iterates, as well, in particular also strong convergence.

Remark 2.10 Let us relate now the results above to the ones given in [58], where accelerated versions of the algorithm from [76] have been proposed. The algorithms in [58] and the ones proposed in this manuscript are designed to solve the same type of problems and under the same hypotheses concerning the operators involved (compare [58, Theorem 3.3] with Theorem 2.2 above and [58, Theorem 3.4] with Theorem 2.3, respectively). The rates of convergence obtained in [58] and here are the same.

On the other hand, our schemes differ from the in [58] in some fundamental aspects. Indeed, we propose here accelerated versions of the algorithm given in [130], which relies on a forward-backward scheme, while in [58] the accelerated versions are with respect to a forward-backward-forward scheme. In contrast to the forward-backward-forward algorithm, which requires additional sequences to be computed, the forward-backward scheme needs fewer steps, thus presents from theoretical point of view an important advantage. This applies also for the accelerated versions of these algorithms. The mentioned advantage is underlined also by the numerical results presented in the subsection 2.1.4. Moreover, one can notice that in Algorithm 2.1 at every iteration when evaluating the operators B_i different step sizes (in form of the parameters $\sigma_{i,n}$) for $i = 1, ..., m$, have been considered, which is not the case with the iterative scheme in [58, Theorem 3.3] where for the evaluation of the same operators the same step size has been used. Individual step sizes possess the advantage that in this way the operators B_i, $i = 1, ..., m$, can be more involved in the algorithm and in the improvement of its convergence properties. A similar remark can be made also for the iterative scheme in [58, Theorem 3.4] and Algorithm 2.2.

2.1.3 Convex optimization problems

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

In order to investigate the applicability of the Algorithm 2.1, we consider the following primal-dual pair of convex optimization problems.

Problem 2.4 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with a η-Lipschitz continuous gradient for $\eta > 0$. Let m be a strictly positive integer and, for every $i \in \{1,..., m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, $g_i \in \Gamma(\mathcal{G}_i)$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. Consider the convex optimization problem*

$$\inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} g_i(L_i x - r_i) + h(x) - \langle x, z\rangle \right\} \tag{2. 31}$$

and its Fenchel-type dual problem

$$\sup_{v_i\in\mathcal{G}_i,\, i=1,..,m} \left\{ -\left(f^* \Box h^*\right)\left(z - \sum_{i=1}^{m} L_i^* v_i\right) - \sum_{i=1}^{m} \left(g_i^*(v_i) + \langle v_i, r_i\rangle\right) \right\}. \tag{2. 32}$$

Considering maximal monotone operators

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

the monotone inclusion problem (2. 5) reads

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in \partial f(\overline{x}) + \sum_{i=1}^{m} L_i^*(\partial g_i(L_i\overline{x} - r_i)) + \nabla h(\overline{x}), \tag{2. 33}$$

while the dual inclusion problem (2. 6) reads

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \left\{ \begin{array}{l} z - \sum_{i=1}^m L_i^* \overline{v}_i \in \partial f(x) + \nabla h(x) \\ \overline{v}_i \in \partial g_i(L_i x - r_i),\ i = 1,..., m. \end{array} \right. \tag{2. 34}$$

If $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to (2. 33)-(2. 34), namely,

$$z - \sum_{i=1}^m L_i^* \overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in \partial g_i(L_i \overline{x} - r_i),\ i = 1,..., m, \tag{2. 35}$$

then $\overline{x}$ is an optimal solution of the problem (2. 31), $(\overline{v}_1,..., \overline{v}_m)$ is an optimal solution of (2. 32) and the optimal objective values of the two problems coincide. Notice that (2. 35) is nothing else than the system of optimality conditions for the primal-dual pair of convex optimization problems (2. 31)-(2. 32).

In case a regularity condition is fulfilled, these optimality conditions are also necessary. More precisely, if the primal problem (2. 31) has an optimal solution $\overline{x}$ and a suitable regularity condition is fulfilled, then there exists $(\overline{v}_1,..., \overline{v}_m)$, an optimal solution to (2. 32), such that $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ satisfies the optimality conditions (2. 35). For the readers convenience, let us present some regularity conditions which are suitable in this context. One of the weakest qualification conditions of interiority-type reads (see, for instance, [76, Proposition 4.3, Remark 4.4])

$$(r_1,..., r_m) \in \operatorname{sqri}\left(\prod_{i=1}^m \operatorname{dom} g_i - \{(L_1 x,..., L_m x) : x \in \operatorname{dom} f\} \right). \tag{2. 36}$$

The condition (2. 36) is fulfilled if one of the following statements holds (see [76, Proposition 4.3]):

(i) $\operatorname{dom} g_i = \mathcal{G}_i,\ i = 1,..., m$;

(ii) $\mathcal{H}$ and $\mathcal{G}_i$ are finite-dimensional and there exists $x \in \operatorname{ri} \operatorname{dom} f$ such that $L_i x - r_i \in \operatorname{ri} \operatorname{dom} g_i,\ i = 1,..., m$.

Another useful and easily verifiable qualification condition guaranteeing that the optimality conditions (2. 35) hold has the following formulation:

(iii) there exists $x' \in \operatorname{dom} f \cap \bigcap_{i=1}^m L_i^{-1}(r_i + \operatorname{dom} g_i)$ such that g_i is continuous at $L_i x' - r_i$, $i = 1,..., m$ (see [38, Remark 2.5] and [52]).

For other qualification conditions for (2. 31)-(2. 32) we refer the reader to consult [26, 37, 38, 40, 131].

The following two statements are particular instances of Algorithm 2.1 and Theorem 2.2, respectively.

Algorithm 2.3

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

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

Theorem 2.4 *Suppose that $f + h$ is γ-strongly convex for some $\gamma > 0$ and let $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ be a primal-dual solution to Problem 2.4. Then the sequences generated by Algorithm 2.3 fulfill for any $n \geq 0$ the inequality*

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

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 $\mathcal{O}(\frac{1}{n})$.

Remark 2.11 Due to the strong convexity of the objective function, the optimization problem (2. 31) in the above theorem has a unique optimal solution (see for example [26, Corollary 11.16]).

Remark 2.12 In case $h(x) = 0$ for all $x \in \mathcal{H}$, one has to choose in Algorithm 2.3 as initial points $\tau_0 > 0, \sigma_{i,0} > 0$, $i = 1, .., m$, with

$$\tau_0 \sum_{i=1}^m \sigma_{i,0}\|L_i\|^2 \leq \sqrt{1 + 2\tau_0\gamma/\lambda}$$

and $\lambda \geq 1$ and to update the sequence $(\theta_n)_{n\geq 0}$ via

$$\theta_n = 1/\sqrt{1 + 2\tau_n\gamma/\lambda}$$

for all $n \geq 0$, in order to obtain a suitable iterative scheme for solving the pair of primal-dual optimization problems (2. 31)-(2. 32) with the same convergence behavior as of Algorithm 2.3. In this situation, when chosing $\lambda = 1$, $m = 1$, $z = 0$ and $r_i = 0$, one obtains an algorithm which is equivalent to the one presented by Chambolle and Pock in [69, Algorithm 2].

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

Problem 2.5 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with a η-Lipschitz continuous gradient for $\eta > 0$. Let m be a strictly positive integer and for every $i \in \{1,..., m\}$ let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that l_i is ν_i^{-1}-strongly convex for $\nu_i > 0$ and $L_i : \mathcal{H} \to \mathcal{G}_i$ a nonzero linear continuous operator. Consider the convex optimization problem*

$$\inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^m (g_i \Box l_i)(L_i x - r_i) + h(x) - \langle x, z\rangle \right\} \tag{2. 37}$$

and its Fenchel-type dual problem

$$\sup_{v_i\in\mathcal{G}_i,\, i=1,..,m} \left\{ -\left(f^* \Box h^*\right)\left(z - \sum_{i=1}^m L_i^* v_i\right) - \sum_{i=1}^m \left(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i\rangle\right) \right\}. \tag{2. 38}$$

Considering the maximal monotone operators

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

according to [26, Proposition 17.10, Theorem 18.15], $D_i^{-1} = \nabla l_i^*$ is a monotone and ν_i-Lipschitz continuous operator for $i = 1,...,m$. The monotone inclusion problem (2. 19) reads

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in \partial f(\overline{x}) + \sum_{i=1}^{m} L_i^*((\partial g_i \Box \partial l_i)(L_i\overline{x} - r_i)) + \nabla h(\overline{x}), \quad (2.\ 39)$$

while the dual inclusion problem (2. 20) reads

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \left\{ \begin{array}{l} z - \sum_{i=1}^m L_i^*\overline{v}_i \in \partial f(x) + \nabla h(x) \\ \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i x - r_i),\ i = 1,...,m. \end{array} \right. \quad (2.\ 40)$$

If $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ is a primal-dual solution to (2. 39)-(2. 40), namely,

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i\overline{x} - r_i),\ i = 1,...,m, \quad (2.\ 41)$$

then $\overline{x}$ is an optimal solution of the problem (2. 37), $(\overline{v}_1,..., \overline{v}_m)$ is an optimal solution of (2. 38) and the optimal objective values of the two problems coincide. Notice that (2. 41) is nothing else than the system of optimality conditions for the primal-dual pair of convex optimization problems (2. 37)-(2. 38).

The assumptions made on l_i guarantees that $g_i \Box l_i \in \Gamma(\mathcal{G}_i)$ (see [26, Corollary 11.16, Proposition 12.14]) and, since $\operatorname{dom}(g_i \Box l_i) = \operatorname{dom} g_i + \operatorname{dom} l_i,\ i = 1,...,m$, one can consider the following qualification condition of interiority-type in order to guarantee (2. 41)

$$(r_1,..., r_m) \in \operatorname{sqri}\left(\prod_{i=1}^{m}(\operatorname{dom} g_i + \operatorname{dom} l_i) - \{(L_1x,..., L_mx) : x \in \operatorname{dom} f\}\right). \quad (2.\ 42)$$

Arguing as above, the condition (2. 42) is fulfilled if one of the following statements holds (see [76, Proposition 4.3])

(i) $\operatorname{dom} g_i + \operatorname{dom} l_i = \mathcal{G}_i$, $i = 1,...,m$;

(ii) $\mathcal{H}$ and $\mathcal{G}_i$ are finite-dimensional and there exists $x \in \operatorname{ri}\operatorname{dom} f$ such that $L_i x - r_i \in \operatorname{ri}\operatorname{dom} g_i + \operatorname{ri}\operatorname{dom} l_i$, $i = 1,...,m$.

The following two statements are particular instances of Algorithm 2.2 and Theorem 2.3, respectively.

Algorithm 2.4

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

For $n \geq 0$ set: $x_{n+1} = \operatorname{prox}_{\tau f}\left[x_n - \tau\left(\sum_{i=1}^m L_i^* v_{i,n} + \nabla h(x_n) - z\right)\right]$
$y_n = x_{n+1} + \theta(x_{n+1} - x_n)$
$v_{i,n+1} = \operatorname{prox}_{\sigma_i g_i^*}[v_{i,n} + \sigma_i(L_i y_n - \nabla l_i^*(v_{i,n}) - r_i)]$, $i = 1,...,m$.

Theorem 2.5 *Suppose that $f + h$ is γ-strongly convex for $\gamma > 0$, $g_i^* + l_i^*$ is δ_i-strongly convex for $\delta_i > 0$, $i = 1,..., m$, and let $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ be a primal-dual solution to Problem 2.5. Then the sequences generated by Algorithm 2.4 fulfill for all $n \geq 0$*

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

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

Remark 2.13 Due to the strong convexity assumptions, the optimization problems (2. 37) and (2. 38) in the above theorem possess unique optimal solutions (see for example [26, Corollary 11.16]).

2.1.4 Numerical experiments

We illustrate the applicability of the theoretical results in the context of two numerical experiments in image processing and pattern recognition in cluster analysis.

Image processing

We compare the numerical performances of Algorithm 2.3 with the ones of other iterative schemes recently introduced in the literature for image denoising. To this end, we treat the nonsmooth regularized convex optimization problem

$$\inf_{x\in\mathbb{R}^k}\left\{\frac{1}{2}\left\|x - b\right\|^2 + \alpha TV(x)\right\}, \tag{2. 43}$$

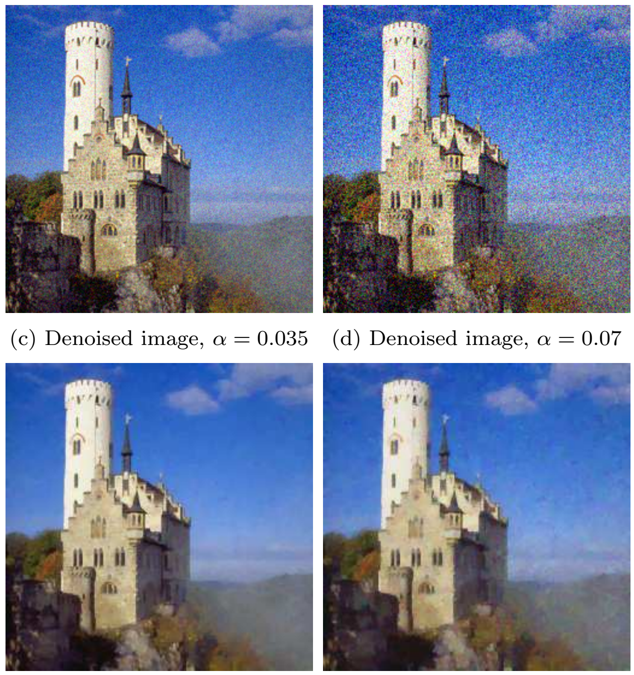
where $TV : \mathbb{R}^k \to \mathbb{R}$ denotes a discrete total variation functional, $\alpha > 0$ is a regularization parameter and $b \in \mathbb{R}^k$ is the observed noisy image. Notice that we consider images of size $k = M \times N$ as vectors $x \in \mathbb{R}^k$, where each pixel denoted by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

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

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

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

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



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

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

Figure 2.1: The noisy images in (a) and (b) were obtained after adding white Gaussian noise with standard deviation $\sigma = 0.06$ and $\sigma = 0.12$, respectively, to the original 256×256 lichtenstein test image. The outputs of Algorithm 2.3 after 100 iterations when solving (2. 43) with isotropic total variation are shown in (c) and (d), respectively.

where in both cases reflexive (Neumann) boundary conditions are assumed. Obviously, in both situations the qualification condition stated in Theorem 2.4 is fulfilled.

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

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

The operator L represents a discretization of the gradient in horizontal and vertical direction. One can easily check that $\|L\|^2 \leq 8$ while for the expression of its adjoint $L^* : \mathcal{Y} \to \mathbb{R}^k$ we refer the reader to [68].

$\varepsilon = 10^{-5}$	isotropic TV		anisotropic TV	
	$\sigma = 0.06$	$\sigma = 0.12$	$\sigma = 0.06$	$\sigma = 0.12$
FB	10.55s (548)	25.78s (1335)	7.83s (517)	12.36s (829)
Algorithm 2.3	3.12s (177)	4.82s (275)	2.66s (202)	3.87s (290)
FBF	19.71s (698)	48.84s (1676)	15.39s (651)	24.60s (1040)
FBF Acc	3.51s (134)	5.94s (208)	3.51s (146)	4.82s (202)
AMA	19.34s (969)	45.94s (2313)	13.58s (901)	22.14s (1448)
AMA Acc	3.38s (132)	5.31s (205)	3.42s (154)	4.80s (230)
Nesterov (dual)	4.48s (146)	6.94s (230)	3.61s (172)	5.42s (249)
FISTA (dual)	3.26s (148)	5.02s (229)	3.14s (173)	4.52s (256)

Table 2.1: Performance evaluation for the images in Figure 2.1. The entries refer, respectively, to the CPU times in seconds and the number of iterations in order to attain a root-mean-square error for the primal iterates below the tolerance level of $\varepsilon = 10^{-5}$.

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

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

and define $\|(y,z)\|_\times = \sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{y_{i,j}^2 + z_{i,j}^2}$. One can check that $\|\cdot\|_\times$ is a norm on $\mathcal{Y}$ and that for every $x \in \mathbb{R}^n$ it holds

$$TV_{\text{iso}}(x) = \|Lx\|_\times. \tag{2. 44}$$

The conjugate function $(\|\cdot\|_\times)^* : \mathcal{Y} \to \overline{\mathbb{R}}$ of $\|\cdot\|_\times$ is for every $(p,q) \in \mathcal{Y}$ given by

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

where

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

Therefore, when considering the *isotropic total variation*, the problem (2. 43) can be formulated as

$$\inf_{x\in\mathbb{R}^k} \{h(x) + g(Lx)\}, \tag{2. 45}$$

where $h : \mathbb{R}^k \to \overline{\mathbb{R}}$,

$$h(x) = \frac{1}{2}\|x-b\|^2$$

is 1-strongly convex with 1-Lipschitz continuous gradient, and $g : \mathcal{Y} \to \mathbb{R}$ is defined as

$$g(u,v) = \alpha\|(u,v)\|_\times.$$

One can show (see [58]) that $g^*(p,q) = \delta_S(p,q)$ for every $(p,q) \in \mathcal{Y}$, where

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

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

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

the projection operator $\text{proj}_S : \mathcal{Y} \to S$ being defined via

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

On the other hand, when considering the *anisotropic total variation*, the problem (2. 43) can be formulated as

$$\inf_{x\in\mathbb{R}^k} \{h(x) + \widetilde{g}(Lx)\}, \tag{2. 46}$$

where the function h is taken as above and $\widetilde{g} : \mathcal{Y} \to \mathbb{R}$ is defined as

$$\widetilde{g}(u,v) = \alpha\|(u,v)\|_1.$$

For every $(p,q) \in \mathcal{Y}$ we have $\widetilde{g}^*(p,q) = \delta_{[-\alpha,\alpha]^k \times [-\alpha,\alpha]^k}(p,q)$ and therefore

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

We consider the *lichtenstein test image* of size 256 times 256 and obtain the corrupted images shown in Figure 2.1 by adding white Gaussian noise with standard deviation $\sigma = 0.06$ and $\sigma = 0.12$, respectively. We then solve (2. 43) by making use of Algorithm 2.3 and by taking into account both instances of the discrete total variation functional. For the picture with noise level $\sigma = 0.06$, we choose the regularization parameter $\alpha = 0.035$, while, in the case when $\sigma = 0.12$, we opted for $\alpha = 0.07$. As initial choices for the parameters occuring in Algorithm 2.3, we let $\gamma = 0.35$, $\eta = 1$, $\lambda = \eta + 1$, $\tau_0 = 0.6\frac{2\gamma}{\eta}$, and $\sigma_0 = \frac{1}{\|L\|^2 \theta_0 \tau_0}$. The reconstructed images after 100 iterations for isotropic total variation are shown in Figure 2.1.

We compare Algorithm 2.3 from the point of view of the CPU time in seconds which is required in order to attain a *root-mean-square error* (RMSE) below the tolerance $\varepsilon = 10^{-5}$ with respect to the primal iterates. Therefore, Table 2.1 shows the achieved results where the comparison is made with the foward-backward method (FB) by Vũ in [130], the foward-backward-forward method (FBF) due to Combettes and Pesquet in [76] and its acceleration (FBF Acc) proposed in [58], the alternating minimization algorithm (AMA) from [128] and its Nesterov type (cf. [105]) acceleration (AMA Acc), as well as the FISTA (cf. [28]) and Nesterov method (cf. [107]), both operating on the dual problem.

As supported by Table 2.1, Algorithm 2.3 competes well against all these methods and provides an accelerated behavior when compared with the forward-backward method by Vũ in Theorem 2.1. In both of these algorithms, we made use of their ability to process the continuously differentiable function $x \mapsto \frac{1}{2}\|x - b\|^2$ via a forward evaluation of its gradient.

Clustering

In cluster analysis one aims for grouping a set of points such that points within the same group are more similar to each other (usually measured via distance functions) than to points in other groups. Clustering can be formulated as a convex optimization problem (see, for instance, [73, 92, 96]). In this example, we consider the minimization problem

$$\inf_{x_i \in \mathbb{R}^n,\, i=1,\ldots,m} \left\{ \frac{1}{2} \sum_{i=1}^{m} \|x_i - u_i\|^2 + \gamma \sum_{i<j} \omega_{ij} \|x_i - x_j\|_p \right\}, \tag{2. 47}$$

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 $i,\, j = 1, \ldots, m$, $i < j$. For each given point $u_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, the variable $x_i \in \mathbb{R}^n$ represents the associated cluster center. Since the objective function is strongly convex, there exists a unique solution to (2. 47).

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

Moreover, the choice of the weights is important as well, since cluster centers may coalesce immediately as γ passes certain critical values. In terms of our weight selection, we use a K-nearest neighbors strategy, as proposed in [73]. Therefore, whenever $i,\, j = 1, ..., m$, $i < j$, we set the weight to $\omega_{ij} = \iota_{ij}^K \exp(-\phi \|x_i - x_j\|_2^2)$, where

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

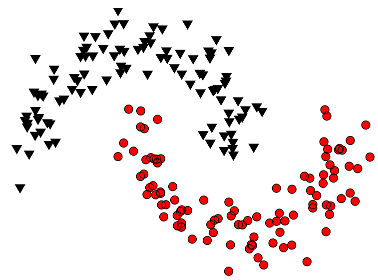


Figure 2.2: Clustering two interlocking half moons. The colors (resp. the shapes) show the correct affiliations.

	$p=2,\ \gamma=5.2$		$p=1,\ \gamma=4$	
	$\varepsilon=10^{-4}$	$\varepsilon=10^{-8}$	$\varepsilon=10^{-4}$	$\varepsilon=10^{-8}$
FB	2.48s (1353)	5.72s (3090)	2.01s (1092)	4.05s (2226)
Algorithm 2.3	2.04s (1102)	4.11s (2205)	1.74s (950)	3.84s (2005)
FBF	7.67s (2123)	17.58s (4879)	6.33s (1781)	13.22s (3716)
FBF Acc	5.05s (1384)	10.27s (2801)	4.83s (1334)	9.98s (2765)
AMA	13.53s (7209)	31.09s (16630)	11.31s (6185)	23.85s (13056)
AMA Acc	3.10s (1639)	15.91s (8163)	2.51s (1392)	12.95s (7148)
Nesterov (dual)	7.85s (3811)	42.69s (21805)	7.46s (3936)	>190s ($>10^5$)
FISTA (dual)	7.55s (4055)	51.01s (27356)	6.55s (3550)	47.81s (26069)

Table 2.2: Performance evaluation for the clustering problem. The entries refer to the CPU times in seconds and the number of iterations, respectively, needed in order to attain a root mean squared error for the iterates below the tolerance ε.

We consider the values $K = 10$ and $\phi = 0.5$, which are the best ones reported in [73] 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 (2. 47) can be equivalently written as

$$\inf_{x\in\mathbb{R}^{mn}} \{h(x) + g(Ax)\}, \tag{2. 48}$$

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^* admit explicit representations.

For our numerical tests we consider the standard dataset consisting of two interlocking half moons in $\mathbb{R}^2$, each of them being composed of 100 points (see Figure 2.2). The stopping criterion asks the root-mean-square error (RMSE) to be less than or equal to a given bound ε which is either $\varepsilon = 10^{-4}$ or $\varepsilon = 10^{-8}$. As tuning parameters we use $\gamma = 4$ for $p = 1$ and $\gamma = 5.2$ for $p = 2$ since both choices lead to a correct separation of the input data into the two half moons.

By taking into consideration the results given in Table 2.2, it shows that Algorithm 2.3 performs slightly better than the forward-backward (FB) method proposed in [130]. One can also see that the acceleration of the forward-backward-forward (FBF) has a positive effect on both CPU times and required iterations compared with the regular method. The alternating minimization algorithm (AMA, cf. [128]) converges slow in this example. Its Nesterov-type acceleration (cf. [105]), however, performs better. The two accelerated first-order methods FISTA (cf. [28]) and the one relying in Nesterov's scheme (cf. [107]), which are both employed on the dual problem, perform surprisingly bad in this case.

2.2 On the convergence rate of a forward-backward type primal-dual splitting algorithm for convex optimization problems

The aim of this section is to investigate the convergence property of the sequence of objective function values of the primal-dual splitting algorithm stated in Theorem 2.1 in the context of convex optimization problems and their Fenchel-type dual. By making use of the so-called restricted primal-dual gap function attached to the problem, we are able to prove a convergence rate of order $\mathcal{O}(1/n)$. The results are formulated in the spirit of the ones given in [69] in a more particular setting.

The starting point of our investigation is the following problem.

Problem 2.6 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with a η^{-1}-Lipschitz continuous gradient for $\eta > 0$. Let m be a strictly positive integer and for $i = 1,...,m$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that l_i is ν_i-strongly convex for $\nu_i > 0$ and $L_i : \mathcal{H} \to \mathcal{G}_i$ a nonzero linear continuous operator. Consider the convex optimization problem*

$$\inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \Box l_i)(L_i x - r_i) + h(x) - \langle x, z\rangle \right\} \tag{2. 49}$$

and its Fenchel-type dual problem

$$\sup_{v_i\in\mathcal{G}_i,\, i=1,..,m} \left\{ -\left(f^* \Box h^*\right) \left(z - \sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} \left(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i\rangle \right) \right\}. \tag{2. 50}$$

The following result is an adaption of [130, Theorem 3.1] to Problem 2.6 to the error-free case and when $\lambda_n = 1$ for all $n \geq 0$.

Theorem 2.6 *(see [130]) In Problem 2.6 suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^{m} L_i^* \left((\partial g_i \Box \partial l_i)(L_i \cdot - r_i) \right) + \nabla h \right). \tag{2. 51}$$

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

$$2 \cdot \min\{\tau^{-1}, \sigma_1^{-1},..., \sigma_m^{-1}\} \cdot \min\{\eta, \nu_1,..., \nu_m\} \cdot \left(1 - \sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2} \right) > 1. \tag{2. 52}$$

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

$$\begin{aligned}
&x_{n+1} = \operatorname{prox}_{\tau f}\left[x_n - \tau\left(\textstyle\sum_{i=1}^{m} L_i^* v_{i,n} + \nabla h(x_n) - z \right) \right]\\
&y_n = 2x_{n+1} - x_n\\
&v_{i,n+1} = \operatorname{prox}_{\sigma_i g_i^*}[v_{i,n} + \sigma_i(L_i y_n - \nabla l_i^*(v_{i,n}) - r_i)],\ i = 1,..., m.
\end{aligned}$$

Then the following statements are true:

(a) there exist $\overline{x} \in \mathcal{H}$, an optimal solution to (2. 49), and $(\overline{v}_1,..., \overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$, an optimal solution to (2. 50), such that the optimal objective values of the two problems coincide, the optimality conditions

$$z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i \overline{x} - r_i),\ i = 1,..., m \tag{2. 53}$$

are fulfilled and $x_n \rightharpoonup \overline{x}$ and $(v_{1,n},..., v_{m,n}) \rightharpoonup (\overline{v}_1,..., \overline{v}_m)$ as $n \to +\infty$;

(b) if h is strongly convex, then $x_n \to \overline{x}$ as $n \to +\infty$;

(c) if l_i^* is strongly convex for some $i \in \{1, ..., m\}$, then $v_{i,n} \to \overline{v}_i$ as $n \to +\infty$.

Before we proceed, some comments are in order.

Remark 2.14 Let us notice that the relation (2. 51) in the above theorem is fulfilled if the primal problem (2. 49) has an optimal solution and the regularity condition (2. 42) holds. Further, let us discuss some conditions ensuring the existence of a primal optimal solution. Suppose that the primal problem (2. 49) is feasible, which means that its optimal objective value is not identical $+\infty$. The existence of optimal solutions for (2. 49) is guaranteed if, for instance, $f + h + \langle \cdot, -z\rangle$ is coercive (that is $\lim_{\|x\|\to\infty}(f + h + \langle \cdot, -z\rangle)(x) = +\infty$) and for all $i = 1, ..., m$, g_i is bounded from below. Indeed, under these circumstances, the objective function of (2. 49) is coercive (use also [26, Corollary 11.16 and Proposition 12.14] to show that for all $i = 1, ..., m$, $g_i \Box l_i$ is bounded from below and $g_i \Box l_i \in \Gamma(\mathcal{G}_i)$) and the statement follows via [26, Corollary 11.15]. On the other hand, if $f + h$ is strongly convex, then the objective function of (2. 49) is strongly convex, too, thus (2. 49) has a unique optimal solution (see [26, Corollary 11.16]).

Remark 2.15 In case $z = 0$, $h \equiv 0$, $r_i = 0$ and $l_i = \delta_{\{0\}}$ for all $i = 1, ..., m$, the optimization problems (2. 49) and (2. 50) become

$$\inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \circ L_i)(x) \right\} \tag{2. 54}$$

and, respectively,

$$\sup_{v_i\in\mathcal{G}_i,\, i=1,..,m} \left\{ -f^*\left(-\sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} g_i^*(v_i) \right\}. \tag{2. 55}$$

It is mentioned in [130, Remark 3.3] that the convergence results in Theorem 2.6 hold if one replaces (2. 52) by the condition

$$\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < 1. \tag{2. 56}$$

The convergence (of an equivalent form) of the algorithm obtained in this setting has been investigated also in [52]. Moreover, the case $m = 1$ has been addressed in [69].

2.2.1 Convergence rate for the objective function values

In the setting of Problem 2.6 we introduce for $B_1 \subseteq \mathcal{H}$ and $B_2 \subseteq \mathcal{G}_1 \times ... \times \mathcal{G}_m$ given nonempty sets the restricted *primal-dual gap function* $\mathcal{G}_{B_1,B_2} \colon \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m \to \overline{\mathbb{R}}$

defined by

$$\begin{aligned}
&\mathcal{G}_{B_1,B_2}(x, v_1, ..., v_m) = \\
&\sup_{(v_1',...,v_m')\in B_2} \left\{ \sum_{i=1}^m \langle L_i x - r_i, v_i'\rangle + f(x) + h(x) - \langle x, z\rangle - \sum_{i=1}^m \Big(g_i^*(v_i') + l_i^*(v_i') \Big) \right\} \\
&- \inf_{x'\in B_1} \left\{ \sum_{i=1}^m \langle L_i x' - r_i, v_i\rangle + f(x') + h(x') - \langle x', z\rangle - \sum_{i=1}^m \Big(g_i^*(v_i) + h_i^*(v_i) \Big) \right\} \\
&= f(x) + h(x) - \langle x, z\rangle \\
&\quad + \sup_{(v_1',...,v_m')\in B_2} \left[\sum_{i=1}^m \langle L_i x - r_i, v_i'\rangle - \sum_{i=1}^m \Big(g_i^*(v_i') + l_i^*(v_i') \Big) \right] \\
&\quad -\Big\{ -\sum_{i=1}^m \big(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i\rangle \big) \\
&\quad + \inf_{x'\in B_1} \Big[\sum_{i=1}^m \langle L_i x', v_i\rangle + f(x') + h(x') - \langle x', z\rangle \Big] \Big\}.
\end{aligned}$$

Remark 2.16 If we consider the primal-dual pair of convex optimization problems from Remark 2.15 in case $m = 1$, then the restricted primal-dual gap function defined above becomes

$$\begin{aligned}
&\mathcal{G}_{B_1,B_2}(x, v_1) = \\
&\sup_{v'\in B_2} \Big\{ \langle Lx - r, v'\rangle + f(x) + h(x) - \langle x, z\rangle - \Big(g^*(v') + l^*(v') \Big) \Big\} \\
&- \inf_{x'\in B_1} \Big\{ \langle Lx' - r, v_1\rangle + f(x') + h(x') - \langle x', z\rangle - \Big(g^*(v_1) + h^*(v_1) \Big) \Big\},
\end{aligned}$$

for $B_1 \subseteq \mathcal{H}$ and $B_2 \subseteq \mathcal{G}_1$, which has been considered in [69].

Remark 2.17 The restricted primal-dual gap function defined above has been used in [58] in order investigate the convergence rate for the sequence of objective function values for the primal-dual splitting algorithm of forward-backward-forward type proposed in [76].

Finally, notice that if $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ satisfies the optimality conditions (2. 53), then $\mathcal{G}_{B_1,B_2}(\overline{x}, \overline{v}_1, ..., \overline{v}_m) \geq 0$ (see also [58, 69]).

We are now able to state the main result of this section.

Theorem 2.7 *In Problem 2.6 suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^m L_i^*\big((\partial g_i \Box \partial l_i)(L_i \cdot - r_i) \big) + \nabla h \right). \tag{2. 57}$$

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

$$\min\{\tau^{-1}, \sigma_1^{-1},..., \sigma_m^{-1}\} \cdot \min\{\eta, \nu_1,..., \nu_m\} \cdot \left(1 - \sqrt{\tau \sum_{i=1}^m \sigma_i \|L_i\|^2} \right) > 1. \tag{2. 58}$$

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

$$\begin{aligned}
&x_{n+1} = \operatorname{prox}_{\tau f}\left[x_n - \tau\big(\textstyle\sum_{i=1}^m L_i^* v_{i,n} + \nabla h(x_n) - z \big) \right] \\
&y_n = 2x_{n+1} - x_n \\
&v_{i,n+1} = \operatorname{prox}_{\sigma_i g_i^*}[v_{i,n} + \sigma_i(L_i y_n - \nabla l_i^*(v_{i,n}) - r_i)], \ i = 1,..., m.
\end{aligned}$$

Then the following statements are true:

(a) *there exist* $\overline{x} \in \mathcal{H}$*, an optimal solution to* (2. 49)*, and* $(\overline{v}_1,...,\overline{v}_m) \in \mathcal{G}_1 \times ... \times \mathcal{G}_m$*, an optimal solution to* (2. 50)*, such that the optimal objective values of the two problems coincide, the optimality conditions*

$$z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i \overline{x} - r_i), \ i = 1,...,m \tag{2. 59}$$

are fulfilled and $x_n \rightharpoonup \overline{x}$ *and* $(v_{1,n},...,v_{m,n}) \rightharpoonup (\overline{v}_1,...,\overline{v}_m)$ *as* $n \rightarrow +\infty$*;*

(b) *if* $B_1 \subseteq \mathcal{H}$ *and* $B_2 \subseteq \mathcal{G}_1 \times ... \times \mathcal{G}_m$ *are nonempty bounded sets, then for* $x^N = \frac{1}{N}\sum_{n=1}^{N} x_{n+1}$ *and* $v_i^N = \frac{1}{N}\sum_{n=1}^{N} v_{i,n}$ *we have* $(x^N, v_1^N, ..., v_m^N) \rightharpoonup (\overline{x}, \overline{v}_1, ..., \overline{v}_m)$ *as* $N \rightarrow +\infty$ *and for all* $N \geq 2$

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

where

$$\begin{aligned} C(B_1, B_2) \ = \ & \sup_{x \in B_1} \left\{ \frac{1}{2\tau}\|x_1 - x\|^2 \right\} + \frac{\sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{2\tau} \|x_1 - x_0\|^2 + \\ & \sup_{(v_1,...,v_m) \in B_2} \left\{ \sum_{i=1}^{m} \frac{1}{2\sigma_i} \|v_{i,0} - v_i\|^2 + \sum_{i=1}^{m} \langle L_i(x_1 - x_0), v_{i,0} - v_i \rangle \right\}; \end{aligned}$$

(c) *if* g_i *is Lipschitz continuous on* $\mathcal{G}_i$ *for every* $i = 1, ..., m$*, then for all* $N \geq 2$ *we have*

$$\begin{aligned} 0 \ \leq \ & \left(f(x^N) + \sum_{i=1}^{m} (g_i \Box l_i)(L_i x^N - r_i) + h(x^N) - \langle x^N, z \rangle \right) \\ & - \left(f(\overline{x}) + \sum_{i=1}^{m} (g_i \Box l_i)(L_i \overline{x} - r_i) + h(\overline{x}) - \langle \overline{x}, z \rangle \right) \\ \leq \ & \frac{C(B_1, B_2)}{N}, \end{aligned} \tag{2. 60}$$

where B_1 *is any bounded and weak sequentially closed set containing the sequence* $(x_n)_{n \in \mathbb{N}}$ *(which is the case if for instance* B_1 *is bounded, convex and closed with respect to the strong topology of* $\mathcal{H}$ *and contains the sequence* $(x_n)_{n \in \mathbb{N}}$*) and* B_2 *is any bounded set containing* $\operatorname{dom} g_1^* \times ... \times \operatorname{dom} g_m^*$*;*

(d) *if* $\operatorname{dom} g_i + \operatorname{dom} l_i = \mathcal{G}_i$ *for every* $i = 1, ..., m$ *and one of the following conditions is fulfilled:*

(d1) $\mathcal{H}$ *is finite-dimensional;*

(d2) $\mathcal{G}_i$ *is finite-dimensional for every* $i = 1, ..., m$*;*

(d3) h *is strongly convex;*

then the inequality (2. 60) *holds for all* $N \geq 2$*, where* B_1 *is taken as in (c) and* B_2 *is any bounded set containing* $\Pi_{i=1}^{m} \cup_{N \geq 2} \partial(g_i \Box l_i)(L_i x^N - r_i)$.

Proof. (a) The statement is a direct consequence of Theorem 2.6, since condition (2. 58) implies (2. 52).

(b) The fact that $(x^N, v_1^N, ..., v_m^N) \rightharpoonup (\overline{x}, \overline{v}_1, ..., \overline{v}_m)$ as $N \rightarrow +\infty$ follows from the Stolz-Cesàro Theorem. Let us show now the inequality concerning the restricted primal-dual gap function. To this aim we fix for the beginning $n \geq 0$.

From the definition of the iterates we derive

$$\frac{1}{\tau}(x_{n+1}-x_{n+2})-\left(\sum_{i=1}^{m}L_i^*v_{i,n+1}+\nabla h(x_{n+1})-z\right)\in\partial f(x_{n+2}),$$

hence the definition of the subdifferential delivers the inequality

$$\begin{aligned} f(x)\geq\quad & f(x_{n+2})+\frac{1}{\tau}\langle x_{n+1}-x_{n+2},x-x_{n+2}\rangle-\left\langle\sum_{i=1}^{m}L_i^*v_{i,n+1}-z,x-x_{n+2}\right\rangle \\ & -\langle\nabla h(x_{n+1}),x-x_{n+2}\rangle\ \forall x\in\mathcal{H}. \end{aligned} \tag{2. 61}$$

Similarly, we deduce

$$\frac{1}{\sigma_i}(v_{i,n}-v_{i,n+1})+L_iy_n-\nabla l_i^*(v_{i,n})-r_i\in\partial g_i^*(v_{i,n+1})\ i=1,...,m,$$

hence for all $i=1,...,m$

$$\begin{aligned} g_i^*(v_i)\ \geq\quad & g_i^*(v_{i,n+1})+\frac{1}{\sigma_i}\langle v_{i,n}-v_{i,n+1},v_i-v_{i,n+1}\rangle+\langle L_iy_n-r_i,v_i-v_{i,n+1}\rangle \\ & -\langle\nabla l_i^*(v_{i,n}),v_i-v_{i,n+1}\rangle\ \forall v_i\in\mathcal{G}_i. \end{aligned} \tag{2. 62}$$

We claim that

$$h(x)-h(x_{n+2})-\langle\nabla h(x_{n+1}),x-x_{n+2}\rangle\geq-\frac{\eta^{-1}}{2}\|x_{n+2}-x_{n+1}\|^2\ \forall x\in\mathcal{H}. \tag{2. 63}$$

Indeed, we have

$$\begin{aligned} & h(x)-h(x_{n+2})-\langle\nabla h(x_{n+1}),x-x_{n+2}\rangle \\ & \quad\geq h(x_{n+1})+\langle\nabla h(x_{n+1}),x-x_{n+1}\rangle-h(x_{n+2})-\langle\nabla h(x_{n+1}),x-x_{n+2}\rangle \\ & \quad= h(x_{n+1})-h(x_{n+2})+\langle\nabla h(x_{n+1}),x_{n+2}-x_{n+1}\rangle \\ & \quad\geq-\frac{\eta^{-1}}{2}\|x_{n+2}-x_{n+1}\|^2, \end{aligned}$$

where the first inequality holds since h is convex and the second one follows follows from Lemma 1.4. Hence (2. 63) holds.

Similarly, one can prove that for all $i=1,...,m$

$$l_i^*(v_i)-l_i^*(v_{i,n+1})-\langle\nabla l_i^*(v_{i,n}),v_i-v_{i,n+1}\rangle\geq-\frac{\nu_i^{-1}}{2}\|v_{i,n+1}-v_{i,n}\|^2\ \forall v_i\in\mathcal{G}_i. \tag{2. 64}$$

By adding the inequalities (2. 61)–(2. 64) and noticing that

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

and

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

we deduce that for all $(x, v_1, ..., v_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$

$$\begin{aligned}
&\frac{\|x_{n+1}-x\|^2}{2\tau} + \sum_{i=1}^m \frac{\|v_{i,n}-v_i\|^2}{2\sigma_i} \\
\geq& \frac{\|x_{n+2}-x\|^2}{2\tau} + \sum_{i=1}^m \frac{\|v_{i,n+1}-v_i\|^2}{2\sigma_i} \\
&+ \frac{1-\eta^{-1}\tau}{2\tau}\|x_{n+1}-x_{n+2}\|^2 + \sum_{i=1}^m \frac{1-\nu_i^{-1}\sigma_i}{2\sigma_i}\|v_{i,n+1}-v_{i,n}\|^2 \\
&+ \left(\sum_{i=1}^m \langle L_i x_{n+2} - r_i, v_i\rangle + f(x_{n+2}) + h(x_{n+2}) - \langle x_{n+2}, z\rangle - \sum_{i=1}^m \Big(g_i^*(v_i) + l_i^*(v_i)\Big)\right) \\
&- \left(\sum_{i=1}^m \langle L_i x - r_i, v_{i,n+1}\rangle + f(x) + h(x) - \langle x, z\rangle - \sum_{i=1}^m \Big(g_i^*(v_{i,n+1}) + l_i^*(v_{i,n+1})\Big)\right) \\
&+ \sum_{i=1}^m \langle L_i(x_{n+2}-y_n), v_{i,n+1}-v_i\rangle.
\end{aligned}$$

Taking into account the definition of y_n, we get the following estimation for the last term:

$$\begin{aligned}
&\langle L_i(x_{n+2}-y_n), v_{i,n+1}-v_i\rangle \\
&=\langle L_i(x_{n+2}-x_{n+1}), v_{i,n+1}-v_i\rangle - \langle L_i(x_{n+1}-x_n), v_{i,n}-v_i\rangle \\
&\quad + \langle L_i(x_{n+1}-x_n), v_{i,n}-v_{i,n+1}\rangle \\
&\geq\langle L_i(x_{n+2}-x_{n+1}), v_{i,n+1}-v_i\rangle - \langle L_i(x_{n+1}-x_n), v_{i,n}-v_i\rangle \\
&\quad - \left(\frac{\sigma_i\|L_i\|^2}{2\sqrt{\tau\sum_{i=1}^m \sigma_i\|L_i\|^2}}\|x_{n+1}-x_n\|^2 + \frac{\sqrt{\tau\sum_{i=1}^m \sigma_i\|L_i\|^2}}{2\sigma_i}\|v_{i,n+1}-v_{i,n}\|^2\right),
\end{aligned}$$

hence we obtain the inequality

$$\begin{aligned}
&\frac{\|x_{n+1}-x\|^2}{2\tau} + \sum_{i=1}^m \frac{\|v_{i,n}-v_i\|^2}{2\sigma_i} \\
\geq& \frac{\|x_{n+2}-x\|^2}{2\tau} + \sum_{i=1}^m \frac{\|v_{i,n+1}-v_i\|^2}{2\sigma_i} \\
&+ \frac{1-\eta^{-1}\tau}{2\tau}\|x_{n+1}-x_{n+2}\|^2 - \frac{\sqrt{\tau\sum_{i=1}^m \sigma_i\|L_i\|^2}}{2\tau}\|x_{n+1}-x_n\|^2 \\
&+ \sum_{i=1}^m \frac{1-\nu_i^{-1}\sigma_i - \sqrt{\tau\sum_{i=1}^m \sigma_i\|L_i\|^2}}{2\sigma_i}\|v_{i,n+1}-v_{i,n}\|^2 \\
&+ \left(\sum_{i=1}^m \langle L_i x_{n+2} - r_i, v_i\rangle + f(x_{n+2}) + h(x_{n+2}) - \langle x_{n+2}, z\rangle - \sum_{i=1}^m \Big(g_i^*(v_i) + l_i^*(v_i)\Big)\right) \\
&- \left(\sum_{i=1}^m \langle L_i x - r_i, v_{i,n+1}\rangle + f(x) + h(x) - \langle x, z\rangle - \sum_{i=1}^m \Big(g_i^*(v_{i,n+1}) + l_i^*(v_{i,n+1})\Big)\right) \\
&+ \sum_{i=1}^m \big(\langle L_i(x_{n+2}-x_{n+1}), v_{i,n+1}-v_i\rangle - \langle L_i(x_{n+1}-x_n), v_{i,n}-v_i\rangle\big).
\end{aligned}$$

Summing up the above inequality from $n = 0$ to $N - 1$, where $N \in \mathbb{N}, N \geq 2$, we get

$$\begin{aligned}
&\frac{\|x_1 - x\|^2}{2\tau} + \sum_{i=1}^{m} \frac{\|v_{i,0} - v_i\|^2}{2\sigma_i} \\
&\geq \frac{\|x_{N+1} - x\|^2}{2\tau} + \sum_{i=1}^{m} \frac{\|v_{i,N} - v_i\|^2}{2\sigma_i} \\
&+ \sum_{n=1}^{N-1} \frac{1 - \eta^{-1}\tau}{2\tau} \|x_{n+1} - x_n\|^2 + \frac{1 - \eta^{-1}\tau}{2\tau} \|x_N - x_{N+1}\|^2 \\
&- \frac{\sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{2\tau} \sum_{n=1}^{N-1} \|x_{n+1} - x_n\|^2 - \frac{\sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{2\tau} \|x_1 - x_0\|^2 \\
&+ \sum_{n=0}^{N-1} \sum_{i=1}^{m} \frac{1 - \nu_i^{-1}\sigma_i - \sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{2\sigma_i} \|v_{i,n+1} - v_{i,n}\|^2 \\
&+ \sum_{n=1}^{N} \left(\sum_{i=1}^{m} \langle L_i x_{n+1} - r_i, v_i \rangle + f(x_{n+1}) + h(x_{n+1}) - \langle x_{n+1}, z \rangle - \sum_{i=1}^{m} \Big(g_i^*(v_i) + l_i^*(v_i) \Big) \right) \\
&- \sum_{n=1}^{N} \left(\sum_{i=1}^{m} \langle L_i x - r_i, v_{i,n} \rangle + f(x) + h(x) - \langle x, z \rangle - \sum_{i=1}^{m} \Big(g_i^*(v_{i,n}) + l_i^*(v_{i,n}) \Big) \right) \\
&+ \sum_{i=1}^{m} \big(\langle L_i(x_{N+1} - x_N), v_{i,N} - v_i \rangle - \langle L_i(x_1 - x_0), v_{i,0} - v_i \rangle \big).
\end{aligned}$$

Further, for the last term we use for all $i = 1, ..., m$ the estimate

$$\begin{aligned}
&\langle L_i(x_{N+1} - x_N), v_{i,N} - v_i \rangle \\
&\quad \geq -\left(\frac{(1 - \eta^{-1}\tau)\sigma_i \|L_i\|^2}{2\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2} \|x_{N+1} - x_N\|^2 + \frac{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}{2\sigma_i (1 - \eta^{-1}\tau)} \|v_{i,N} - v_i\|^2 \right)
\end{aligned}$$

(notice that $1 - \eta^{-1}\tau > 0$ due to (2. 58)) and conclude that for all $(x, v_1, ..., v_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$

$$\begin{aligned}
&\frac{\|x_{N+1} - x\|^2}{2\tau} + \sum_{i=1}^{m} \frac{1 - \eta^{-1}\tau - \tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}{2\sigma_i} \|v_{i,N} - v_i\|^2 \\
&+ \sum_{n=1}^{N-1} \frac{1 - \eta^{-1}\tau - \sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{2\tau} \|x_{n+1} - x_n\|^2 \\
&+ \sum_{n=0}^{N-1} \sum_{i=1}^{m} \frac{1 - \nu_i^{-1}\sigma_i - \sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{2\sigma_i} \|v_{i,n+1} - v_{i,n}\|^2 \\
&+ \sum_{n=1}^{N} \left(\sum_{i=1}^{m} \langle L_i x_{n+1} - r_i, v_i \rangle + f(x_{n+1}) + h(x_{n+1}) - \langle x_{n+1}, z \rangle - \sum_{i=1}^{m} \Big(g_i^*(v_i) + l_i^*(v_i) \Big) \right) \\
&- \sum_{n=1}^{N} \left(\sum_{i=1}^{m} \langle L_i x - r_i, v_{i,n} \rangle + f(x) + h(x) - \langle x, z \rangle - \sum_{i=1}^{m} \Big(g_i^*(v_{i,n}) + l_i^*(v_{i,n}) \Big) \right) \\
&\leq \frac{\|x_1 - x\|^2}{2\tau} + \sum_{i=1}^{m} \frac{\|v_{i,0} - v_i\|^2}{2\sigma_i} \\
&+ \frac{\sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{2\tau} \|x_1 - x_0\|^2 + \sum_{i=1}^{m} \langle L_i(x_1 - x_0), v_{i,0} - v_i \rangle.
\end{aligned}$$

We can discard the first four terms in the left-hand side of the above inequality, since due to (2. 58) we have

$$1-\eta^{-1}\tau-\sqrt{\tau\sum_{i=1}^{m}\sigma_i\|L_i\|^2}>0 \tag{2. 65}$$

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

$$1-\nu_i^{-1}\sigma_i-\sqrt{\tau\sum_{i=1}^{m}\sigma_i\|L_i\|^2}>0. \tag{2. 66}$$

Thus we obtain for all $(x,v_1,...,v_m)\in\mathcal{H}\times\mathcal{G}_1\times...\times\mathcal{G}_m$ that

$$\sum_{n=1}^{N}\left(\sum_{i=1}^{m}\langle L_ix_{n+1}-r_i,v_i\rangle+f(x_{n+1})+h(x_{n+1})-\langle x_{n+1},z\rangle-\sum_{i=1}^{m}\Big(g_i^*(v_i)+l_i^*(v_i)\Big)\right)$$
$$-\sum_{n=1}^{N}\left(\sum_{i=1}^{m}\langle L_ix-r_i,v_{i,n}\rangle+f(x)+h(x)-\langle x,z\rangle-\sum_{i=1}^{m}\Big(g_i^*(v_{i,n})+l_i^*(v_{i,n})\Big)\right)$$
$$\leq\frac{\|x_1-x\|^2}{2\tau}+\sum_{i=1}^{m}\frac{\|v_{i,0}-v_i\|^2}{2\sigma_i}+\frac{\sqrt{\tau\sum_{i=1}^{m}\sigma_i\|L_i\|^2}}{2\tau}\|x_1-x_0\|^2+\sum_{i=1}^{m}\langle L_i(x_1-x_0),v_{i,0}-v_i\rangle.$$

The conclusion follows by passing into the previous inequality to the supremum over $x\in B_1$ and $(v_1,...,v_m)\in B_2$ and by taking into account the definition of $(x^N,v_i^N,...,v_m^N)$ and the convexity of the functions f,h and g_i^*,h_i^*, $i=1,...,m$.

(c) According to [37, Proposition 4.4.6], the set $\operatorname{dom} g_i^*$ is bounded for $i=1,...,m$. Since B_1 is weak sequentially closed and $x_n\rightharpoonup\overline{x}$, we have $\overline{x}\in B_1$.

Let be $N\geq 2$ fixed. We get from (b) that

$$\frac{C(B_1,B_2)}{N}\geq\mathcal{G}(x^N,v_1^N,...,v_m^N)\geq$$
$$f(x^N)+h(x^N)-\langle x^N,z\rangle+\sum_{i=1}^{m}\sup_{v_i'\in\operatorname{dom} g_i^*}\left\{\langle L_ix^N-r_i,v_i'\rangle-(g_i^*(v_i')+l_i^*(v_i')\right\}$$
$$-\left(\sum_{i=1}^{m}\langle L_i\overline{x}-r_i,v_i^N\rangle+f(\overline{x})+h(\overline{x})-\langle\overline{x},z\rangle-\sum_{i=1}^{m}\left(g_i^*(v_i^N)+l_i^*(v_i^N)\right)\right).$$

Further, since $\operatorname{dom} l_i^*=\mathcal{G}_i$ for $i=1,...,m$, it follows

$$\sup_{v_i'\in\operatorname{dom} g_i^*}\left\{\langle L_ix^N-r_i,v_i'\rangle-(g_i^*(v_i')+l_i^*(v_i')\right\}$$
$$=\sup_{v_i'\in\operatorname{dom} g_i^*\cap\operatorname{dom} l_i^*}\left\{\langle L_ix^N-r_i,v_i'\rangle-(g_i^*(v_i')+l_i^*(v_i')\right\}$$
$$=(g_i^*+l_i^*)^*(L_ix^N-r_i)=(g_i^{**}\Box l_i^{**})(L_ix^N-r_i)=(g_i\Box l_i)(L_ix^N-r_i),$$

where we used [26, Proposition 15.2] and the celebrated Fenchel-Moreau Theorem (see for example [26, Theorem 13.32]). Furthermore, the Young-Fenchel inequality (see [26, Proposition 13.13]) guarantees that for all $i=1,...,m$

$$g_i^*(v_i^N)+l_i^*(v_i^N)-\langle L_i\overline{x}-r_i,v_i^N\rangle=(g_i\Box h_i)^*(v_i^N)-\langle L_i\overline{x}-r_i,v_i^N\rangle\geq-(g_i\Box l_i)(L_i\overline{x}-r_i)$$

and the conclusion follows.

(d) We notice first that each of the conditions (d1),(d2) and (d3) implies that

$$L_i x^N \to L_i \overline{x} \text{ as } N \to +\infty \text{ for all } i = 1, ..., m. \tag{2. 67}$$

Indeed, in case of (d1) we use that $x^N \rightharpoonup \overline{x}$ as $N \to +\infty$, in case (d2) that $L_i x^N \rightharpoonup L_i \overline{x}$ as $N \to +\infty$ (which is a consequence of $x^N \rightharpoonup \overline{x}$ as $N \to +\infty$), while in the last case we appeal Theorem 2.1(b).

We fix $i \in \{1, ..., m\}$ and show first that $\cup_{N\geq 1} \partial(g_i \Box l_i)(L_i x^N - r_i)$ is a nonempty bounded set. The function $g_i \Box l_i$ belongs to $\Gamma(\mathcal{H})$, as already mentioned in Remark 2.14. Further, as $\operatorname{dom}(g_i \Box l_i) = \operatorname{dom} g_i + \operatorname{dom} l_i = \mathcal{G}_i$, it follows that $g_i \Box l_i$ is everywhere continuous (see [26, Corollary 8.30]) and, consequently, everywhere subdifferentiable (see [26, Proposition 16.14(iv)]). Hence, the claim concerning the nonemptiness of the set $\cup_{N\geq 1} \partial(g_i \Box l_i)(L_i x^N - r_i)$ is true. Moreover, since the subdifferential of $g_i \Box l_i$ is locally bounded at $L_i \overline{x} - r_i$ (see [26, Proposition 16.14(iii)]) and $L_i x^N - r_i \to L_i \overline{x} - r_i$ as $N \to +\infty$ we easily derive from [26, Proposition 16.14(iii) and (ii)] that the set $\cup_{N\geq 1} \partial(g_i \Box l_i)(L_i x^N - r_i)$ is bounded.

Now we prove that the inequality (2. 60) holds. Similarly as in (c), we have

$$\frac{C(B_1, B_2)}{N} \geq \mathcal{G}(x^N, v_1^N, ..., v_m^N) \geq$$

$$f(x^N) + h(x^N) - \langle x^N, z \rangle$$

$$+ \sum_{i=1}^{m} \sup_{v_i' \in \cup_{N'\geq 2} \partial(g_i \Box l_i)(L_i x^{N'} - r_i)} \left\{ \langle L_i x^N - r_i, v_i' \rangle - (g_i^*(v_i') + l_i^*(v_i') \right\}$$

$$- \left(\sum_{i=1}^{m} \langle L_i \overline{x} - r_i, v_i^N \rangle + f(\overline{x}) + h(\overline{x}) - \langle \overline{x}, z \rangle - \sum_{i=1}^{m} \left(g_i^*(v_i^N) + l_i^*(v_i^N) \right) \right).$$

Further, for all $i = 1, ..., m$ and for all $N \geq 1$ we have

$$\sup_{v_i' \in \cup_{N'\geq 2} \partial(g_i \Box l_i)(L_i x^{N'} - r_i)} \left\{ \langle L_i x^N - r_i, v_i' \rangle - (g_i^*(v_i') + l_i^*(v_i') \right\}$$

$$\geq \sup_{v_i' \in \partial(g_i \Box l_i)(L_i x^N - r_i)} \left\{ \langle L_i x^N - r_i, v_i' \rangle - (g_i^*(v_i') + l_i^*(v_i') \right\}$$

$$= (g_i \Box l_i)(L_i x^N - r_i),$$

where the last equality follows since $\partial(g_i \Box l_i)(L_i x^N - r_i) \neq \emptyset$ via

$$\langle L_i x^N - r_i, v_i' \rangle - (g_i^*(v_i') + l_i^*(v_i')) = \langle L_i x^N - r_i, v_i' \rangle - (g_i \Box h_i)^*(v_i') = (g_i \Box l_i)(L_i x^N - r_i),$$

which holds for every $v_i' \in \partial(g_i \Box l_i)(L_i x^N - r_i)$ (see [26, Proposition 16.9]).

Using the same arguments as at the end of the proof of statement (c), the conclusion follows. □

Remark 2.18 When considering the particular instance as described in Remark 2.15 with the additional assumption $m = 1$, similar results to Theorem 2.7 have been reported in [69] for an equivalent form of the algorithm.

Remark 2.19 The conclusion of the above theorem remains true if condition (2. 58) is replaced by (2. 52), (2. 65) and (2. 66). Moreover, if one works in the setting of Remark 2.15, one can show that the conclusion of Theorem 2.7 remains valid if instead of (2. 58) one assumes (2. 56).

Remark 2.20 Let us mention that in Theorem 2.7(c) and (d) one can chose for B_1 any bounded set containing $\overline{x}$.

Remark 2.21 If f is Lipschitz continuous, then, similarly to Theorem 2.7(c), one can prove via Theorem 2.7(b) a convergence rate of order $\mathcal{O}(1/n)$ for the sequence of values of the objective function of the dual problem (2. 38). The same conclusion follows in case f has full domain and one of the conditions (d1), (d2) and (d3') is fulfilled, where (d3') assumes that l_i^* is strongly convex for any $i = 1, ..., m$.

Remark 2.22 For more recent advances concerning convergence rates for the objective function values via primal-dual splitting techniques we invite the reader to consult [81, 82], where also nonergodic convergence results are reported and also to [72], where a multi-step acceleration scheme in the sense of Nesterov is incorporated into the primal-dual method in order to increase the speed of convergence.

2.2.2 Numerical experiments

We illustrate the theoretical results obtained in the previous subsection by means of a problem occurring in imaging. For the applications discussed in this section the images have been normalized in order to make their pixels range in the closed interval from 0 to 1.

TV-based image deblurring

The considered numerical experiment addresses an ill-conditioned linear inverse problem which arises in image deblurring. For a given matrix $A \in \mathbb{R}^{n\times n}$ describing a blur operator and a given vector $b \in \mathbb{R}^n$ representing the blurred and noisy image, the task is to estimate the unknown original image $\overline{x} \in \mathbb{R}^n$ fulfilling

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

To this end we solve the following regularized convex minimization problem

$$\inf_{x\in[0,1]^n} \left\{ \|Ax - b\|_1 + \lambda(TV_{\text{iso}}(x) + \|x\|^2) \right\}, \tag{2. 68}$$

where $\lambda > 0$ is a regularization parameter and $TV_{\text{iso}} : \mathbb{R}^n \to \mathbb{R}$ is the discrete isotropic total variation functional. 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 $i = 1, \ldots, M$ and $j = 1, \ldots, N$.

We invite the reader to consult the section corresponding to the numerical experiments in chapter 2 for the definition of the isotropic total variational functional. By using also the notations specified there, the optimization problem (2. 68) can be written in the form of

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

where

$$f : \mathbb{R}^n \to \overline{\mathbb{R}},\ f(x) = \delta_{[0,1]^n}(x),$$
$$g_1 : \mathbb{R}^n \to \mathbb{R},\ g_1(y) = \|y - b\|_1,$$
$$g_2 : \mathcal{Y} \to \mathbb{R},\ g_2(y, z) = \lambda \|(y, z)\|_\times$$

and

$$h : \mathbb{R}^n \to \mathbb{R},\ h(x) = \lambda\|x\|^2$$

(notice that the functions l_i are taken to be $\delta_{\{0\}}$ for $i = 1, 2$). For every $p \in \mathbb{R}^n$, it holds

$$g_1^*(p) = \delta_{[-1,1]^n}(p) + p^T b,$$

(a) Original image (b) Blurred and noisy image (c) Reconstructed image

Figure 2.3: Figure (a) shows the original 256×256 boat test image, figure (b) shows the blurred and noisy image and figure (c) shows the averaged iterate generated by the algorithm after 400 iterations.

while for every $(p, q) \in \mathcal{Y}$, we have

$$g_2^*(p, q) = \delta_S(p, q),$$

with $S = \{(p, q) \in \mathcal{Y} : \|(p, q)\|_{\times *} \leq \lambda\}$. Moreover, h is differentiable with $\eta^{-1} := 2\lambda$-Lipschitz continuous gradient. We solved this problem by the algorithm considered in Theorem 2.7 above and to this end we made use of the following formulae

$$\operatorname{prox}_{\gamma f}(x) = \operatorname{proj}_{[0,1]^n}(x) \ \ \forall x \in \mathbb{R}^n$$

$$\operatorname{prox}_{\gamma g_1^*}(p) = \operatorname{proj}_{[-1,1]^n}(p - \gamma b) \ \ \forall p \in \mathbb{R}^n$$

$$\operatorname{prox}_{\gamma g_2^*}(p, q) = \operatorname{proj}_S(p, q) \ \ \forall (p, q) \in \mathcal{Y},$$

where $\gamma > 0$ and the projection operator $\operatorname{proj}_S : \mathcal{Y} \to S$ is defined as (see [58])

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

For the experiments we considered the 256×256 boat test image and constructed the blurred image by making use of a Gaussian blur operator of size 9×9 and standard deviation 4. In order to obtain the blurred and noisy image we added a zero-mean white Gaussian noise with standard deviation 10^{-3}. Figure 2.3 shows the original boat test image and the blurred and noisy one. It also shows the image reconstructed by the algorithm after 400 iterations in terms of the averaged iterate, when taking as regularization parameter $\lambda = 0.001$ and by choosing as parameters $\sigma_1 = 0.01, \sigma_2 = 0.7, \tau = 0.49$. On the other hand, in Figure 2.4 a comparison of the decrease of the objective function values is provided, in terms of the last and averaged iterates, underlying the rate of convergence of order $\mathcal{O}(1/n)$ for the latter.

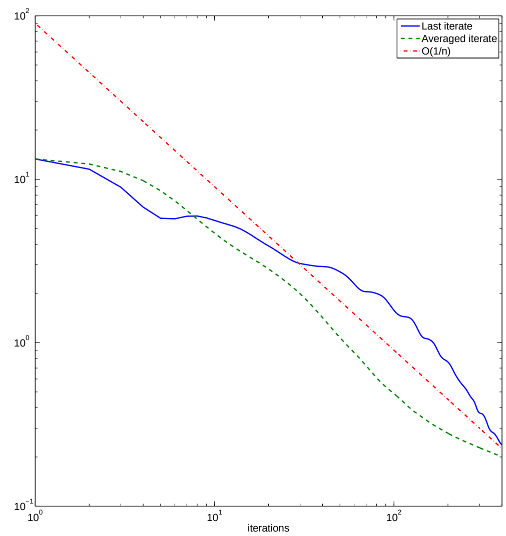


Figure 2.4: The figure shows the relative error in terms of function values for both the last and the averaged iterate generated by the algorithm after 400 iterations.

Chapter 3

Splitting algorithms involving inertial terms

In this chapter we introduce and investigate several inertial-type proximal-splitting algorithms designed for solving highly structured monotone inclusion problems. In Section 3.1 we propose an inertial version of the forward-backward-forward proximal splitting algorithm, while in Section 3.2 the attention is focused on primal-dual algorithms of Douglas-Rachford-type. In Section 3.3 we formulate and investigate an inertial forward-backward algorithm in the context of solving nonconvex optimization problems with analytic futures.

The following convergence results will be used in the proof of the main results in this chapter. These statements can be seen as generalizations of Lemma 1.2 and turn out to be useful for proving the first property in the Opial Lemma.

Lemma 3.1 *(see [3–5]) 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 statements 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* $\lim_{n\to+\infty} \varphi_n = \varphi^*$.

An easy consequence of Lemma 3.1 is the following result.

Lemma 3.2 *Let $(\varphi_n)_{n\in\mathbb{N}}, (\delta_n)_{n\in\mathbb{N}}, (\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be sequences in $[0,+\infty)$ such that $\varphi_{n+1} \leq -\beta_n + \varphi_n + \alpha_n(\varphi_n - \varphi_{n-1}) + \delta_n$ 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 statements 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* $\lim_{n\to+\infty} \varphi_n = \varphi^*$*;*

(iii) $\sum_{n\in\mathbb{N}} \beta_n < +\infty$.

3.1 Tseng's type inertial primal-dual algorithms for monotone inclusions

In this section we propose an inertial forward-backward-forward-type proximal splitting algorithms associated to a monotone inclusion problem. An essential argument in the favor of Tseng's type splitting algorithms is given by the fact that they can be used when solving a larger class of monotone inclusion problems. This is for instance of importance when considering primal-dual splitting methods, as shown by the approach described in [62].

3.1.1 An inertial forward-backward-forward splitting algorithm

This section is dedicated to the formulation of an inertial forward-backward-forward splitting algorithm which approaches the set of zeros of the sum of two maximally monotone operators, one of them being single-valued and Lipschitz continuous, and to the investigation of its convergence properties.

Theorem 3.1 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ a monotone and β-Lipschitz continuous operator for some $\beta > 0$. Suppose that $\operatorname{zer}(A+B) \neq \emptyset$ and consider the following iterative scheme:*

$$(\forall n \geq 1) \quad \left\{ \begin{array}{l} p_n = J_{\lambda_n A}[x_n - \lambda_n B x_n + \alpha_{1,n}(x_n - x_{n-1})] \\ x_{n+1} = p_n + \lambda_n(Bx_n - Bp_n) + \alpha_{2,n}(x_n - x_{n-1}), \end{array} \right.$$

where x_0 and x_1 are arbitrarily chosen in $\mathcal{H}$. Consider $\lambda, \sigma > 0$ and $\alpha_1, \alpha_2 \geq 0$ such that

$$12\alpha_2^2 + 9(\alpha_1+\alpha_2) + 4\sigma < 1 \text{ and } \lambda \leq \lambda_n \leq \frac{1}{\beta}\sqrt{\frac{1 - 12\alpha_2^2 - 9(\alpha_1+\alpha_2) - 4\sigma}{12\alpha_2^2 + 8(\alpha_1+\alpha_2) + 4\sigma + 2}} \ \forall n \geq 1 \tag{3. 1}$$

and for $i = 1, 2$ the nondecreasing sequences $(\alpha_{i,n})_{n\geq 1}$ fulfilling

$$0 \leq \alpha_{i,n} \leq \alpha_i \ \forall n \geq 1.$$

Then there exists $\overline{x} \in \operatorname{zer}(A+B)$ such that the following statements are true:

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

(b) $x_n \rightharpoonup \overline{x}$ and $p_n \rightharpoonup \overline{x}$ as $n \to +\infty$;

(c) Suppose that one of the following conditions is satisfied:

(i) $A + B$ is demiregular at $\overline{x}$;

(ii) A or B is uniformly monotone at $\overline{x}$.

Then $x_n \to \overline{x}$ and $p_n \to \overline{x}$ as $n \to +\infty$.

Proof. Let z be a fixed element in $\operatorname{zer}(A+B)$, that is $-Bz \in Az$, and $n \geq 1$. From the definition of the resolvent we deduce

$$\frac{1}{\lambda_n}(x_n - p_n) - Bx_n + \frac{\alpha_{1,n}}{\lambda_n}(x_n - x_{n-1}) \in Ap_n.$$

Further, taking into account the relation between p_n and x_{n+1} in the algorithm, we obtain

$$\frac{1}{\lambda_n}(x_n - x_{n+1}) - Bp_n + \frac{\alpha_{1,n}+\alpha_{2,n}}{\lambda_n}(x_n - x_{n-1}) \in Ap_n. \tag{3. 2}$$

The monotonicity of A delivers the inequality

$$0 \leq \left\langle \frac{1}{\lambda_n}(x_n - x_{n+1}) - Bp_n + \frac{\alpha_{1,n}+\alpha_{2,n}}{\lambda_n}(x_n - x_{n-1}) + Bz, p_n - z \right\rangle,$$

hence

$$0 \leq \frac{1}{\lambda_n}\langle x_n - x_{n+1}, p_n - z\rangle + \langle Bz - Bp_n, p_n - z\rangle + \frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}\langle x_n - x_{n-1}, p_n - z\rangle. \tag{3. 3}$$

Since B is monotone, we have

$$\langle Bz - Bp_n, p_n - z\rangle \leq 0.$$

Moreover,

$$\begin{aligned}\langle x_n - x_{n+1}, p_n - z\rangle &= \langle x_n - x_{n+1}, p_n - x_{n+1}\rangle + \langle x_n - x_{n+1}, x_{n+1} - z\rangle \\ &= \frac{\|x_n - x_{n+1}\|^2}{2} + \frac{\|p_n - x_{n+1}\|^2}{2} - \frac{\|x_n - p_n\|^2}{2} \\ &\quad + \frac{\|x_n - z\|^2}{2} - \frac{\|x_n - x_{n+1}\|^2}{2} - \frac{\|x_{n+1} - z\|^2}{2}.\end{aligned}$$

In a similar way we obtain

$$\begin{aligned}\langle x_n - x_{n-1}, p_n - z\rangle &= \langle x_n - x_{n-1}, x_n - z\rangle + \langle x_n - x_{n-1}, p_n - x_n\rangle \\ &= \frac{\|x_n - x_{n-1}\|^2}{2} + \frac{\|x_n - z\|^2}{2} - \frac{\|x_{n-1} - z\|^2}{2} \\ &\quad + \frac{\|p_n - x_{n-1}\|^2}{2} - \frac{\|x_n - x_{n-1}\|^2}{2} - \frac{\|x_n - p_n\|^2}{2}.\end{aligned}$$

Further, by using that B is β-Lipschitz continuous, we have

$$\|x_{n+1} - p_n\|^2 \leq 2\lambda_n^2\beta^2\|x_n - p_n\|^2 + 2\alpha_{2,n}^2\|x_n - x_{n-1}\|^2$$

and

$$\|p_n - x_{n-1}\|^2 \leq 2\|x_n - p_n\|^2 + 2\|x_n - x_{n-1}\|^2.$$

The above estimates together with (3. 3) imply

$$\begin{aligned}0 \leq &\left(\frac{1}{2\lambda_n} + \frac{\alpha_{1,n} + \alpha_{2,n}}{2\lambda_n}\right)\|x_n - z\|^2 \\ &- \frac{1}{2\lambda_n}\|x_{n+1} - z\|^2 - \frac{\alpha_{1,n} + \alpha_{2,n}}{2\lambda_n}\|x_{n-1} - z\|^2 \\ &+ \left(\lambda_n\beta^2 - \frac{1}{2\lambda_n} + \frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n} - \frac{\alpha_{1,n} + \alpha_{2,n}}{2\lambda_n}\right)\|x_n - p_n\|^2 \\ &+ \left(\frac{\alpha_{2,n}^2}{\lambda_n} + \frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}\right)\|x_n - x_{n-1}\|^2,\end{aligned}$$

from which we further obtain, after multiplying with $2\lambda_n$,

$$\begin{aligned}\|x_{n+1} - z\|^2 - (1 + \alpha_{1,n} + \alpha_{2,n})\|x_n - z\|^2 + (\alpha_{1,n} + \alpha_{2,n})\|x_{n-1} - z\|^2 \\ \leq -(1 - \alpha_{1,n} - \alpha_{2,n} - 2\lambda_n^2\beta^2)\|x_n - p_n\|^2 \\ + 2(\alpha_{2,n}^2 + \alpha_{1,n} + \alpha_{2,n})\|x_n - x_{n-1}\|^2.\end{aligned} \tag{3. 4}$$

By using the bounds given for the sequences $(\lambda_n)_{n\geq 1}$, $(\alpha_{1,n})_{n\geq 1}$ and $(\alpha_{2,n})_{n\geq 1}$, one can easily show by taking into account (3. 1) that

$$2\lambda_n^2\beta^2 < 1 - \alpha_1 - \alpha_2 \leq 1 - \alpha_{1,n} - \alpha_{2,n},$$

thus

$$1 - \alpha_{1,n} - \alpha_{2,n} - 2\lambda_n^2\beta^2 > 0.$$

Furthermore, since

$$\begin{aligned}\|x_{n+1} - x_n\|^2 &= \|p_n - x_n + \lambda_n(Bx_n - Bp_n) + \alpha_{2,n}(x_n - x_{n-1})\|^2 \\ &\leq 2(1+\lambda_n\beta)^2\|x_n - p_n\|^2 + 2\alpha_{2,n}^2\|x_n - x_{n-1}\|^2,\end{aligned}$$

we obtain from (3. 4)

$$\begin{aligned}&\|x_{n+1} - z\|^2 - (1 + \alpha_{1,n} + \alpha_{2,n})\|x_n - z\|^2 + (\alpha_{1,n} + \alpha_{2,n})\|x_{n-1} - z\|^2 \\ &\leq -\frac{1 - \alpha_{1,n} - \alpha_{2,n} - 2\lambda_n^2\beta^2}{2(1+\lambda_n\beta)^2}\|x_{n+1} - x_n\|^2 + \gamma_n\|x_n - x_{n-1}\|^2,\end{aligned} \tag{3. 5}$$

where

$$\gamma_n := 2(\alpha_{2,n}^2 + \alpha_{1,n} + \alpha_{2,n}) + \frac{\alpha_{2,n}^2(1 - \alpha_{1,n} - \alpha_{2,n} - 2\lambda_n^2\beta^2)}{(1+\lambda_n\beta)^2} > 0.$$

(a) For the proof of this statement we are going to use some techniques from [5]. We define the sequences

$$\varphi_n := \|x_n - z\|^2 \ \forall n \in \mathbb{N}$$

and

$$\mu_n := \varphi_n - (\alpha_{1,n} + \alpha_{2,n})\varphi_{n-1} + \gamma_n\|x_n - x_{n-1}\|^2 \ \forall n \geq 1.$$

Using the monotonicity of $(\alpha_{i,n})_{n\geq 1}$, $i = 1, 2$, and the fact that $\varphi_n \geq 0$ for all $n \in \mathbb{N}$, we get

$$\begin{aligned}\mu_{n+1} - \mu_n \leq\ & \varphi_{n+1} - (1 + \alpha_{1,n} + \alpha_{2,n})\varphi_n + (\alpha_{1,n} + \alpha_{2,n})\varphi_{n-1} \\ &+ \gamma_{n+1}\|x_{n+1} - x_n\|^2 - \gamma_n\|x_n - x_{n-1}\|^2,\end{aligned}$$

which gives by (3. 5)

$$\mu_{n+1} - \mu_n \leq -\left(\frac{1 - \alpha_{1,n} - \alpha_{2,n} - 2\lambda_n^2\beta^2}{2(1+\lambda_n\beta)^2} - \gamma_{n+1}\right)\|x_{n+1} - x_n\|^2 \ \forall n \geq 1. \tag{3. 6}$$

We claim that

$$\frac{1 - \alpha_{1,n} - \alpha_{2,n} - 2\lambda_n^2\beta^2}{2(1+\lambda_n\beta)^2} - \gamma_{n+1} \geq \sigma \ \forall n \geq 1. \tag{3. 7}$$

Indeed, this follows by taking into account that for all $n \geq 1$

$$\begin{aligned}&\alpha_{1,n} + \alpha_{2,n} + 2(\lambda_n\beta)^2 + 2(1+\lambda_n\beta)^2(\gamma_{n+1} + \sigma) \\ &\leq \alpha_1 + \alpha_2 + 2(\lambda_n\beta)^2 + 2(1+\lambda_n\beta)^2(3\alpha_2^2 + 2(\alpha_1 + \alpha_2) + \sigma) \\ &\leq \alpha_1 + \alpha_2 + 2(\lambda_n\beta)^2 + 4(1 + (\lambda_n\beta)^2)(3\alpha_2^2 + 2(\alpha_1 + \alpha_2) + \sigma) \\ &\leq 1.\end{aligned}$$

In the above estimates we used the upper bounds for $(\alpha_{i,n})_{n\geq 1}$, $i = 1, 2$, that

$$\gamma_{n+1} \leq 2(\alpha_2^2 + \alpha_1 + \alpha_2) + \alpha_2^2 \ \forall n \in \mathbb{N}$$

and the assumptions in (3. 1).

We obtain from (3. 6) and (3. 7) that

$$\mu_{n+1} - \mu_n \leq -\sigma\|x_{n+1} - x_n\|^2 \ \forall n \geq 1. \tag{3. 8}$$

Hence, the sequence $(\mu_n)_{n\geq 1}$ is nonincreasing and so, we can let $M \geq 0$ be an upper bound of it, that is $\mu_n \leq M$ for all $n \geq 1$. The bounds for $(\alpha_{i,n})_{n\geq 1}$, $i = 1, 2$, deliver

$$-(\alpha_1 + \alpha_2)\varphi_{n-1} \leq \varphi_n - (\alpha_1 + \alpha_2)\varphi_{n-1} \leq \mu_n \leq M \ \forall n \geq 1. \tag{3. 9}$$

We obtain for all $n \geq 1$

$$\begin{aligned}\varphi_n &\leq (\alpha_1 + \alpha_2)^n \varphi_0 + M \sum_{k=0}^{n-1} (\alpha_1 + \alpha_2)^k \\ &\leq (\alpha_1 + \alpha_2)^n \varphi_0 + \frac{M}{1 - \alpha_1 - \alpha_2}.\end{aligned}$$

Combining (3. 8) and (3. 9) we have for all $n \geq 1$

$$\begin{aligned}\sigma \sum_{k=1}^{n} \|x_{k+1} - x_k\|^2 \leq \ & \mu_1 - \mu_{n+1} \\ &\leq \mu_1 + (\alpha_1 + \alpha_2)\varphi_n \\ &\leq \mu_1 + (\alpha_1 + \alpha_2)^{n+1}\varphi_0 + \frac{M(\alpha_1 + \alpha_2)}{1 - \alpha_1 - \alpha_2},\end{aligned}$$

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

Combining this relation with (3. 4) and Lemma 3.2 it yields

$$\sum_{n\geq 1} \left(1 - \alpha_{1,n} - \alpha_{2,n} - 2\lambda_n^2\beta^2\right) \|x_n - p_n\|^2 < +\infty.$$

Moreover, from (3. 7) we have $1 - \alpha_{1,n} - \alpha_{2,n} - 2\lambda_n^2\beta^2 \geq 2\sigma(1 + \lambda\beta)^2$ for all $n \geq 1$ and obtain, consequently, $\sum_{n\geq 1} \|x_n - p_n\|^2 < +\infty$.

(b) We are going to use Lemma 1.1 for the proof of this statement. We proved above that for an arbitrary $z \in \operatorname{zer}(A + B)$ the inequality (3. 4) is true. By part (a) and Lemma 3.2 it follows that $\lim_{n\to+\infty} \|x_n - z\|$ exists. On the other hand, let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, that is, let be the subsequence $(x_{n_k})_{k\in\mathbb{N}}$ fulfilling $x_{n_k} \rightharpoonup x$ as $k \to +\infty$. Since $x_n - p_n \to 0$ as $n \to +\infty$, we get $p_{n_k} \rightharpoonup x$ as $k \to +\infty$. Since $A+B$ is maximally monotone (see [26, Corollary 20.25 and Corollary 24.4]), its graph is sequentially closed in the weak-strong topology of $\mathcal{H} \times \mathcal{H}$ (see [26, Proposition 20.33(ii)]). As $(\lambda_n)_{n\geq 1}$ and $(\alpha_{i,n})_{n\geq 1}$, $i = 1, 2$, are bounded, we derive from (3. 2) and part (a) that $0 \in (A+B)x$, hence $x \in \operatorname{zer}(A+B)$. By Lemma 1.1 there exists $\overline{x} \in \operatorname{zer}(A + B)$ such that $x_n \rightharpoonup \overline{x}$ as $n \to +\infty$. In view of (a) we have $p_n \rightharpoonup \overline{x}$ as $n \to +\infty$.

(c) Since (ii) implies that $A + B$ is uniformly monotone at $\overline{x}$, hence demiregular at $\overline{x}$, it is sufficient to prove the statement under condition (i). Since $p_n \rightharpoonup \overline{x}$ and $\frac{1}{\lambda_n}(x_n - x_{n+1}) + \frac{\alpha_{1,n}+\alpha_{2,n}}{\lambda_n}(x_n - x_{n-1}) \to 0$ as $n \to +\infty$, the result follows easily from (3. 2) and the definition of demiregular operators. □

Remark 3.1 Assuming that $\alpha_2 = 0$, which enforces $\alpha_{2,n} = 0$ for all $n \geq 1$, the conclusions of Theorem 3.1 remains valid if one takes as upper bound for $(\lambda_n)_{n\geq 1}$ the expression $\frac{1}{\beta}\sqrt{\frac{1-5\alpha_1-2\sigma}{4\alpha_1+2\sigma+1}}$. This is due to the fact that in this situation one can use in its proof the improved inequalities $\|x_{n+1} - p_n\|^2 \leq \lambda_n^2\beta^2\|x_n - p_n\|^2$ and $\|x_{n+1} - x_n\|^2 \leq (1 + \lambda_n\beta)^2\|x_n - p_n\|^2$ for all $n \geq 1$. On the other hand, let us also notice that the algorithmic scheme obtained in this way and its convergence properties can be seen as generalizations of the corresponding statements given for the error-free case of the classical forward-backward-forward algorithm proposed by Tseng in [129] (see also [62, Theorem 2.5]). Indeed, if we further set $\alpha_1 = 0$, having as consequence that $\alpha_{1,n} = 0$ for all $n \geq 1$, we obtain nothing else than the iterative scheme from [62, 129]. Notice that for $\varepsilon \in (0, 1/(\beta+1))$, one can chose $\lambda := \varepsilon$ and $\sigma := \frac{1-(1-\varepsilon)^2}{2(1+(1-\varepsilon)^2)}$. In this case the sequence $(\lambda_n)_{n\geq 1}$ must fulfill the inequalities $\varepsilon \leq \lambda_n \leq \frac{1}{\beta}\sqrt{\frac{1-2\sigma}{2\sigma+1}} = \frac{1-\varepsilon}{\beta}$ for all $n \geq 1$, which is exactly the situation considered in [62].

Remark 3.2 In case $Bx = 0$ for all $x \in \mathcal{H}$ the iterative scheme in Theorem 3.1 becomes

$$x_{n+1} = J_{\lambda_n A}[x_n + \alpha_{1,n}(x_n - x_{n-1})] + \alpha_{2,n}(x_n - x_{n-1}) \ \forall n \geq 1,$$

and is to the best of our knowledge new and can be regarded as an extension of the classical proximal-point algorithm (see [122]) in the context of solving the monotone inclusion problem $0 \in Ax$. If, additionally, $\alpha_2 = 0$, which enforces as already noticed $\alpha_{2,n} = 0$ for all $n \geq 1$, we get the algorithm

$$x_{n+1} = J_{\lambda_n A}[x_n + \alpha_{1,n}(x_n - x_{n-1})],$$

the convergence of which has been investigated in [5].

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

In this section we employ the inertial forward-backward-forward splitting algorithm proposed above to the concomitantly solving of a primal monotone inclusion problem involving mixtures of linearly composed and parallel-sum type operators and its Attouch-Théra-type dual problem. We consider the following setting.

Problem 3.1 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ a monotone and μ-Lipschitz continuous operator for $\mu > 0$. Let m be a strictly positive integer and, for every $i \in \{1,..., m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone operator, let $D_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be monotone such that D_i^{-1} is ν_i-Lipschitz continuous for $\nu_i > 0$ and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. The problem is to solve the primal inclusion*

$$\textit{find } \overline{x} \in \mathcal{H} \textit{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^*\big((B_i \Box D_i)(L_i\overline{x} - r_i)\big) + C\overline{x} \tag{3. 10}$$

together with the dual inclusion

$$\textit{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \textit{ such that } \exists x \in \mathcal{H} : \left\{ \begin{array}{l} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in Ax + Cx \\ \overline{v}_i \in (B_i \Box D_i)(L_i x - r_i),\ i = 1,..., m. \end{array} \right. \tag{3. 11}$$

Similar to Problem 2.1, one can define primal-dual solutions to Problem 3.1.

Problem 3.1 covers a large class of monotone inclusion problems and we refer the reader to consult [76] for several interesting particular instances of it. The main result of this section follows.

Theorem 3.2 *In Problem 3.1 suppose that*

$$z \in \operatorname{ran}\left(A + \sum_{i=1}^{m} L_i^*\big((B_i \Box D_i)(L_i \cdot - r_i)\big) + C\right). \tag{3. 12}$$

Chose $x_0, x_1 \in \mathcal{H}$ and $v_{i,0}, v_{i,1} \in \mathcal{G}_i$, $i = 1, ..., m$, and set

$$(\forall n \geq 1) \quad \left\{ \begin{array}{rl} p_{1,n} = & J_{\lambda_n A}[x_n - \lambda_n(Cx_n + \sum_{i=1}^m L_i^* v_{i,n} - z) + \alpha_{1,n}(x_n - x_{n-1})] \\ p_{2,i,n} = & J_{\lambda_n B_i^{-1}}[v_{i,n} + \lambda_n(L_i x_n - D_i^{-1} v_{i,n} - r_i) + \alpha_{1,n}(v_{i,n} - v_{i,n-1})], \\ & i = 1, ..., m \\ v_{i,n+1} = & \lambda_n L_i(p_{1,n} - x_n) + \lambda_n(D_i^{-1} v_{i,n} - D_i^{-1} p_{2,i,n}) + p_{2,i,n} \\ & +\alpha_{2,n}(v_{i,n} - v_{i,n-1}), i = 1, ..., m \\ x_{n+1} = & \lambda_n \sum_{i=1}^m L_i^*(v_{i,n} - p_{2,i,n}) + \lambda_n(Cx_n - Cp_{1,n}) + p_{1,n} \\ & +\alpha_{2,n}(x_n - x_{n-1}). \end{array} \right.$$

Consider $\lambda, \sigma > 0$ and $\alpha_1, \alpha_2 \geq 0$ such that

$$12\alpha_2^2 + 9(\alpha_1 + \alpha_2) + 4\sigma < 1 \text{ and } \lambda \leq \lambda_n \leq \frac{1}{\beta}\sqrt{\frac{1 - 12\alpha_2^2 - 9(\alpha_1 + \alpha_2) - 4\sigma}{12\alpha_2^2 + 8(\alpha_1 + \alpha_2) + 4\sigma + 2}} \ \forall n \geq 1,$$

where

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

and for $i = 1, 2$ the nondecreasing sequences $(\alpha_{i,n})_{n \geq 1}$ fulfilling

$$0 \leq \alpha_{i,n} \leq \alpha_i \ \forall n \geq 1.$$

Then the following statements are true:

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

(b) There exists $(\overline{x}, \overline{v}_1, ..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ a primal-dual solution to Problem 3.1 such that the following hold:

(i) $x_n \rightharpoonup \overline{x}$, $p_{1,n} \rightharpoonup \overline{x}$ and, for $i = 1, ..., m$, $v_{i,n} \rightharpoonup \overline{v}_i$ and $p_{2,i,n} \rightharpoonup \overline{v}_i$ as $n \to +\infty$;

(ii) If $A + C$ is uniformly monotone at $\overline{x}$, then $x_n \to \overline{x}$ and $p_{1,n} \to \overline{x}$ as $n \to +\infty$.

(iii) If $B_i^{-1} + D_i^{-1}$ is uniformly monotone at $\overline{v}_i$ for some $i \in \{1, ..., m\}$, then $v_{i,n} \to \overline{v}_i$ and $p_{2,i,n} \to \overline{v}_i$ as $n \to +\infty$.

Proof. We apply Theorem 3.1 in an appropriate product space and make use to this end of a construction similar to the one considered in [76]. We endow the product space $\boldsymbol{\mathcal{K}} = \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ with the inner product and the associated norm defined for all $(x, v_1, ..., v_m), (y, w_1, ..., w_m) \in \boldsymbol{\mathcal{K}}$ as

$$\langle (x, v_1, ..., v_m), (y, w_1, ..., w_m) \rangle_{\boldsymbol{\mathcal{K}}} = \langle x, y \rangle_{\mathcal{H}} + \sum_{i=1}^m \langle v_i, w_i \rangle_{\mathcal{G}_i}$$

and

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

respectively.

We introduce the operators $\boldsymbol{M} : \boldsymbol{\mathcal{K}} \rightrightarrows \boldsymbol{\mathcal{K}}$,

$$\boldsymbol{M}(x, v_1, ..., v_m) = (-z + Ax) \times (r_1 + B_1^{-1} v_1) \times .... \times (r_m + B_m^{-1} v_m)$$

and $\boldsymbol{Q} : \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}}$,

$$\boldsymbol{Q}(x, v_1, ..., v_m) = \left(Cx + \sum_{i=1}^{m} L_i^* v_i, -L_1 x + D_1^{-1} v_1, ..., -L_m x + D_m^{-1} v_m\right)$$

and show that Theorem 3.1 can be applied for the operators $\boldsymbol{M}$ and $\boldsymbol{Q}$ in the product space $\boldsymbol{\mathcal{K}}$. Let us start by noticing that

$$(3.\ 12) \Leftrightarrow \operatorname{zer}(\boldsymbol{M} + \boldsymbol{Q}) \neq \emptyset$$

and

$$(x, v_1, ..., v_m) \in \operatorname{zer}(\boldsymbol{M}+\boldsymbol{Q}) \Leftrightarrow (x, v_1, ..., v_m) \text{ is a primal-dual solution of Problem 3.1.} \tag{3. 13}$$

Further, since A and B_i, $i = 1, ..., m$, are maximally monotone, $\boldsymbol{M}$ is maximally monotone, too (see [26, Propositions 20.22 and 20.23]). On the other hand, $\boldsymbol{Q}$ is monotone and β-Lipschitz continuous (see, for instance, the proof of [76, Theorem 3.1]).

For every $(x, v_1, ..., v_m) \in \boldsymbol{\mathcal{K}}$ and every $\lambda > 0$ we have (see [26, Proposition 23.16])

$$J_{\lambda \boldsymbol{M}}(x, v_1, ..., v_m) = (J_{\lambda A}(x + \lambda z), J_{\lambda B_1^{-1}}(v_1 - \lambda r_1), ..., J_{\lambda B_m^{-1}}(v_m - \lambda r_m)).$$

Set

$$\boldsymbol{x}_n = (x_n, v_{1,n}, ..., v_{m,n}) \ \forall n \in \mathbb{N} \text{ and } \boldsymbol{p}_n = (p_{1,n}, p_{2,1,n}, ..., p_{2,m,n}) \ \forall n \geq 1.$$

In the light of the above considerations it follows that the iterative scheme in the statement of Theorem 3.2 can be equivalently written as

$$(\forall n \geq 1) \quad \left\{ \begin{array}{l} \boldsymbol{p}_n = J_{\lambda_n \boldsymbol{M}}[\boldsymbol{x}_n - \lambda_n \boldsymbol{Q}\boldsymbol{x}_n + \alpha_{1,n}(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})] \\ \boldsymbol{x}_{n+1} = \boldsymbol{p}_n + \lambda_n(\boldsymbol{Q}\boldsymbol{x}_n - \boldsymbol{Q}p_n) + \alpha_{2,n}(\boldsymbol{x}_n - \boldsymbol{x}_{n-1}), \end{array} \right.$$

which is nothing else than the algorithm stated in Theorem 3.1 formulated for the operators $\boldsymbol{M}$ and $\boldsymbol{Q}$.

(a) Is a direct consequence of Theorem 3.1(a).

(b)(i) Is a direct consequence of Theorem 3.1(b) and (3. 13).

(b)(ii) Let $n \geq 1$ be fixed. From the definition of the resolvent we get

$$\frac{1}{\lambda_n}(x_n - p_{1,n}) - Cx_n - \sum_{i=1}^{m} L_i^* v_{i,n} + z + \frac{\alpha_{1,n}}{\lambda_n}(x_n - x_{n-1}) \in Ap_{1,n}.$$

The update rule for x_n yields

$$\frac{1}{\lambda_n}(p_{1,n} - x_{n+1}) + Cx_n + \sum_{i=1}^{m} L_i^*(v_{i,n} - p_{2,i,n}) + \frac{\alpha_{2,n}}{\lambda_n}(x_n - x_{n-1}) = Cp_{1,n},$$

hence,

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

Further, since $z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in (A + C)\overline{x}$ and $A + C$ is uniformly monotone at $\overline{x}$, there exists an increasing function $\phi_{A,C} : [0, +\infty) \to [0, +\infty]$ that vanishes only at 0, such that

$$\Big\langle p_{1,n} - \overline{x}, \frac{1}{\lambda_n}(x_n - x_{n+1}) - \sum_{i=1}^m L_i^* p_{2,i,n} + z$$
$$+\frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}(x_n - x_{n-1}) - \Big(z - \sum_{i=1}^m L_i^* \overline{v}_i\Big)\Big\rangle$$
$$\geq \phi_{A,C}(\|p_{1,n} - \overline{x}\|),$$

thus

$$\frac{1}{\lambda_n}\langle p_{1,n} - \overline{x}, x_n - x_{n+1}\rangle + \left\langle p_{1,n} - \overline{x}, \sum_{i=1}^m L_i^*(\overline{v}_i - p_{2,i,n})\right\rangle$$
$$+\frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}\langle p_{1,n} - \overline{x}, x_n - x_{n-1}\rangle \geq \phi_{A,C}(\|p_{1,n} - \overline{x}\|). \qquad (3.\ 14)$$

In a similar way, for $i = 1, ..., m$, the definition of $p_{2,i,n}$ yields

$$\frac{1}{\lambda_n}(v_{i,n} - p_{2,i,n}) + L_i x_n - D_i^{-1} v_{i,n} - r_i + \frac{\alpha_{1,n}}{\lambda_n}(v_{i,n} - v_{i,n-1}) \in B_i^{-1} p_{2,i,n}$$

and from

$$\frac{1}{\lambda_n}(p_{2,i,n} - v_{i,n+1}) + L_i p_{1,n} - L_i x_n + D_i^{-1} v_{i,n} + \frac{\alpha_{2,n}}{\lambda_n}(v_{i,n} - v_{i,n-1}) = D_i^{-1} p_{2,i,n}$$

we further obtain

$$\frac{1}{\lambda_n}(v_{i,n} - v_{i,n+1}) + L_i p_{1,n} - r_i + \frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}(v_{i,n} - v_{i,n-1}) \in (B_i^{-1} + D_i^{-1}) p_{2,i,n}.$$

Moreover, since $L_i\overline{x} - r_i \in (B_i^{-1} + D_i^{-1})\overline{v}_i$, the monotonicity of $B_i^{-1} + D_i^{-1}, i = 1, ..., m$, yields the inequality

$$\Big\langle \frac{1}{\lambda_n}(v_{i,n} - v_{i,n+1}) + L_i p_{1,n} - r_i$$
$$+\frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}(v_{i,n} - v_{i,n-1}) - (L_i\overline{x} - r_i), p_{2,i,n} - \overline{v}_i\Big\rangle$$
$$\geq 0$$

hence

$$\frac{1}{\lambda_n}\sum_{i=1}^m\langle v_{i,n} - v_{i,n+1}, p_{2,i,n} - \overline{v}_i\rangle + \left\langle p_{1,n} - \overline{x}, \sum_{i=1}^m L_i^*(p_{2,i,n} - \overline{v}_i)\right\rangle$$
$$+\frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}\sum_{i=1}^m\langle v_{i,n} - v_{i,n-1}, p_{2,i,n} - \overline{v}_i\rangle \geq 0. \qquad (3.\ 15)$$

Summing up the inequalities (3. 14) and (3. 15) we obtain for all $n \geq 1$

$$\frac{1}{\lambda_n}\langle p_{1,n} - \overline{x}, x_n - x_{n+1}\rangle + \frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}\langle p_{1,n} - \overline{x}, x_n - x_{n-1}\rangle$$
$$+\frac{1}{\lambda_n}\sum_{i=1}^m\langle v_{i,n} - v_{i,n+1}, p_{2,i,n} - \overline{v}_i\rangle + \frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}\sum_{i=1}^m\langle v_{i,n} - v_{i,n-1}, p_{2,i,n} - \overline{v}_i\rangle$$
$$\geq \phi_{A,C}(\|p_{1,n} - \overline{x}\|). \qquad (3.\ 16)$$

It then follows from (a), (b)(i) and the boundedness of the sequences $(\alpha_{i,n})_{n\geq 1}$, $i = 1, 2$ and $(\lambda_n)_{n\geq 1}$ that $\lim_{n\to+\infty} \phi_{A,C}(\|p_{1,n}-\overline{x}\|) = 0$, thus $p_{1,n} \to \overline{x}$ as $n \to +\infty$. From (a) we get that $x_n \to \overline{x}$ as $n \to +\infty$.

(b)(iii) In this case one can show that instead of (3. 16) one has for all $n \geq 1$

$$\begin{aligned}
&\frac{1}{\lambda_n}\langle p_{1,n} - \overline{x}, x_n - x_{n+1}\rangle + \frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}\langle p_{1,n} - \overline{x}, x_n - x_{n-1}\rangle \\
&+\frac{1}{\lambda_n}\sum_{j=1}^{m}\langle v_{j,n} - v_{j,n+1}, p_{2,j,n} - \overline{v}_j\rangle + \frac{\alpha_{1,n} + \alpha_{2,n}}{\lambda_n}\sum_{j=1}^{m}\langle v_{j,n} - v_{j,n-1}, p_{2,j,n} - \overline{v}_j\rangle \\
&\geq \phi_{B_i^{-1},D_i^{-1}}(\|p_{2,i,n} - \overline{v}_i\|).
\end{aligned} \tag{3. 17}$$

where $\phi_{B_i^{-1},D_i^{-1}} : [0, +\infty) \to [0, +\infty]$ is an increasing function that vanishes only at 0. The same arguments as in (b)(ii) provide the desired conclusion. □

Remark 3.3 The case $\alpha_1 = \alpha_2 = 0$, which enforces $\alpha_{1,n} = \alpha_{2,n} = 0$ for all $n \geq 1$, shows that the error-free case of the forward-backward-forward algorithm considered in [76, Theorem 3.1] is a particular case of the iterative scheme introduced in Theorem 3.2. We refer to Remark 3.1 for a discussion on how to choose the parameters λ and σ in order to get exactly the bounds from [76, Theorem 3.1].

3.1.3 Convex optimization problems

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

Problem 3.2 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with a μ-Lipschitz continuous gradient for $\mu > 0$. Let m be a strictly positive integer and for all $i \in \{1,..., m\}$ let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that l_i is ν_i^{-1}-strongly convex for $\nu_i > 0$ and $L_i : \mathcal{H} \to \mathcal{G}_i$ a nonzero linear continuous operator. Consider the convex optimization problem*

$$\inf_{x\in\mathcal{H}}\left\{f(x) + \sum_{i=1}^{m}(g_i\Box l_i)(L_i x - r_i) + h(x) - \langle x, z\rangle\right\} \tag{3. 18}$$

and its Fenchel-type dual problem

$$\sup_{v_i\in\mathcal{G}_i,\, i=1,..,m}\left\{-\left(f^*\Box h^*\right)\left(z - \sum_{i=1}^{m} L_i^* v_i\right) - \sum_{i=1}^{m}\left(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i\rangle\right)\right\}. \tag{3. 19}$$

Considering the maximal monotone operators

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

according to [26, Proposition 17.10, Theorem 18.15], $D_i^{-1} = \nabla l_i^*$ is a monotone and ν_i-Lipschitz continuous operator for $i = 1,..., m$. The monotone inclusion problem (3. 10) reads

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in \partial f(\overline{x}) + \sum_{i=1}^{m} L_i^*((\partial g_i\Box\partial l_i)(L_i\overline{x} - r_i)) + \nabla h(\overline{x}), \tag{3. 20}$$

while the dual inclusion problem (3. 11) reads

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } \exists x \in \mathcal{H} : \left\{\begin{array}{l} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in \partial f(x) + \nabla h(x) \\ \overline{v}_i \in (\partial g_i\Box\partial l_i)(L_i x - r_i),\ i = 1,..., m. \end{array}\right. \tag{3. 21}$$

The optimality conditions concerning the primal-dual pair of optimization problems (3. 18)-(3. 19) are nothing else than

$$z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i \overline{x} - r_i),\ i = 1,...,m. \quad (3.\ 22)$$

Notice that the aforementioned optimality conditions are also necessary in case the regularity condition (2. 42) is fulfilled.

The following statement is a particular instance of Theorem 3.2.

Theorem 3.3 *In Problem 3.2 suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^{m} L_i^*\big((\partial g_i \Box \partial l_i)(L_i \cdot - r_i)\big) + \nabla h\right). \quad (3.\ 23)$$

Chose $x_0, x_1 \in \mathcal{H}$ *and* $v_{i,0}, v_{i,1} \in \mathcal{G}_i,\ i = 1, ..., m,$ *and set*

$$(\forall n \geq 1) \quad \left\{ \begin{array}{rl} p_{1,n} = & \operatorname{prox}_{\lambda_n f}[x_n - \lambda_n(\nabla f(x_n) + \sum_{i=1}^{m} L_i^* v_{i,n} - z) + \alpha_{1,n}(x_n - x_{n-1})] \\ p_{2,i,n} = & \operatorname{prox}_{\lambda_n g_i^*}[v_{i,n} + \lambda_n(L_i x_n - \nabla l_i^*(v_{i,n}) - r_i) + \alpha_{1,n}(v_{i,n} - v_{i,n-1})], \\ & i = 1, ..., m \\ v_{i,n+1} = & \lambda_n L_i(p_{1,n} - x_n) + \lambda_n(\nabla l_i^*(v_{i,n}) - \nabla l_i^*(p_{2,i,n})) + p_{2,i,n} \\ & +\alpha_{2,n}(v_{i,n} - v_{i,n-1}), i = 1, ..., m \\ x_{n+1} = & \lambda_n \sum_{i=1}^{m} L_i^*(v_{i,n} - p_{2,i,n}) + \lambda_n(\nabla h(x_n) - \nabla h(p_{1,n})) + p_{1,n} \\ & +\alpha_{2,n}(x_n - x_{n-1}). \end{array} \right.$$

Consider $\lambda, \sigma > 0$ *and* $\alpha_1 \geq 0, \alpha_2 \geq 0$ *such that*

$$12\alpha_2^2 + 9(\alpha_1 + \alpha_2) + 4\sigma < 1 \text{ and } \lambda \leq \lambda_n \leq \frac{1}{\beta}\sqrt{\frac{1 - 12\alpha_2^2 - 9(\alpha_1 + \alpha_2) - 4\sigma}{12\alpha_2^2 + 8(\alpha_1 + \alpha_2) + 4\sigma + 2}}\ \forall n \geq 1,$$

where

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

and for $i = 1, 2$ *the nondecreasing sequences* $(\alpha_{i,n})_{n\geq 1}$ *fulfilling*

$$0 \leq \alpha_{i,n} \leq \alpha_i\ \forall n \geq 1.$$

Then the following statements are true:

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

(b) There exists $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ *satisfying the optimality conditions* (3. 22), *hence* $\overline{x}$ *is an optimal solution of the problem* (3. 18), $(\overline{v}_1,..., \overline{v}_m)$ *is an optimal solution of* (3. 19) *and the optimal objective values of the two problems coincide, such that the following hold:*

  (i) $x_n \rightharpoonup \overline{x}$, $p_{1,n} \rightharpoonup \overline{x}$ *and, for* $i = 1, ..., m$, $v_{i,n} \rightharpoonup \overline{v}_i$ *and* $p_{2,i,n} \rightharpoonup \overline{v}_i$ *as* $n \to +\infty$;

  (ii) If $f + h$ *is uniformly convex, then* $x_n \to \overline{x}$ *and* $p_{1,n} \to \overline{x}$ *as* $n \to +\infty$;

  (iii) If $g_i^* + l_i^*$ *is uniformly convex for some* $i \in \{1, ..., m\}$, *then* $v_{i,n} \to \overline{v}_i$ *and* $p_{2,i,n} \to \overline{v}_i$ *as* $n \to +\infty$.

Remark 3.4 Under the hypotheses considered in Problem 3.2, condition (3. 23) is fulfilled if the primal problem (3. 18) has an optimal solution and the regularity condition (2. 42) holds.

3.2 Inertial Douglas–Rachford splitting for monotone inclusions

The aim of this section is the investigation of an inertial-type Douglas-Rachford algorithm for solving monotone inclusion problems and the illustration of the numerical advantages in comparison with its noninertial version.

3.2.1 An inertial Douglas–Rachford splitting algorithm

This subsection 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 [99], 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 3.4 *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:*

$$x_{n+1} = x_n+\alpha_n(x_n-x_{n-1})+\lambda_n\Big[T\big(x_n+\alpha_n(x_n-x_{n-1})\big)-x_n-\alpha_n(x_n-x_{n-1})\Big] \ \forall n \geq 1 \tag{3. 24}$$

where x_0, x_1 are arbitrarily chosen in M, $(\alpha_n)_{n\geq 1}$ is nondecreasing and fulfills

$$0 \leq \alpha_n \leq \alpha < 1 \ \forall n \geq 1$$

and $\lambda, \sigma, \delta > 0$ are such that

$$\delta > \frac{\alpha^2(1+\alpha)+\alpha\sigma}{1-\alpha^2} \text{ and } 0 < \lambda \leq \lambda_n \leq \frac{\delta - \alpha\Big[\alpha(1+\alpha)+\alpha\delta+\sigma\Big]}{\delta\Big[1+\alpha(1+\alpha)+\alpha\delta+\sigma\Big]} \ \forall n \geq 1. \tag{3. 25}$$

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. We 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 $n \geq 1$:

$$x_{n+1} = w_n + \lambda_n(Tw_n - w_n). \tag{3. 26}$$

We fix an element $y \in \operatorname{Fix} T$ and $n \geq 1$. It follows from (1. 28) and the nonexpansiveness of T that

$$\|x_{n+1} - y\|^2 = (1-\lambda_n)\|w_n - y\|^2 + \lambda_n\|Tw_n - Ty\|^2 - \lambda_n(1-\lambda_n)\|Tw_n - w_n\|^2$$
$$\leq \|w_n - y\|^2 - \lambda_n(1-\lambda_n)\|Tw_n - w_n\|^2. \qquad (3.\ 27)$$

Applying (1. 28) 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 (3. 27) we obtain

$$\|x_{n+1} - y\|^2 - (1+\alpha_n)\|x_n - y\|^2 + \alpha_n\|x_{n-1} - y\|^2 \leq -\lambda_n(1-\lambda_n)\|Tw_n - w_n\|^2$$
$$+ \alpha_n(1+\alpha_n)\|x_n - x_{n-1}\|^2. \qquad (3.\ 28)$$

Furthermore, we have

$$\|Tw_n - w_n\|^2 = \left\|\frac{1}{\lambda_n}(x_{n+1} - x_n) + \frac{\alpha_n}{\lambda_n}(x_{n-1} - x_n)\right\|^2$$
$$= \frac{1}{\lambda_n^2}\|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2}{\lambda_n^2}\|x_n - x_{n-1}\|^2 + 2\frac{\alpha_n}{\lambda_n^2}\langle x_{n+1} - x_n, x_{n-1} - x_n\rangle$$
$$\geq \frac{1}{\lambda_n^2}\|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2}{\lambda_n^2}\|x_n - x_{n-1}\|^2$$
$$+ \frac{\alpha_n}{\lambda_n^2}\left(-\rho_n\|x_{n+1} - x_n\|^2 - \frac{1}{\rho_n}\|x_n - x_{n-1}\|^2\right), \qquad (3.\ 29)$$

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

We derive from (3. 28) and (3. 29) the inequality

$$\|x_{n+1} - y\|^2 - (1+\alpha_n)\|x_n - y\|^2 + \alpha_n\|x_{n-1} - y\|^2$$
$$\leq \frac{(1-\lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n}\|x_{n+1} - x_n\|^2 + \gamma_n\|x_n - x_{n-1}\|^2, \qquad (3.\ 30)$$

where

$$\gamma_n := \alpha_n(1+\alpha_n) + \alpha_n(1-\lambda_n)\frac{1-\rho_n\alpha_n}{\rho_n\lambda_n} > 0, \qquad (3.\ 31)$$

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 (3. 31), it follows

$$\gamma_n = \alpha_n(1+\alpha_n) + \alpha_n(1-\lambda_n)\delta \leq \alpha(1+\alpha) + \alpha\delta \ \forall n \geq 1. \qquad (3.\ 32)$$

In the following we use some techniques from [5] adapted to our setting. We define the sequences

$$\varphi_n := \|x_n - y\|^2 \ \forall n \in \mathbb{N}$$

and

$$\mu_n := \varphi_n - \alpha_n\varphi_{n-1} + \gamma_n\|x_n - x_{n-1}\|^2 \ \forall n \geq 1.$$

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 (3. 30), we have

$$\mu_{n+1} - \mu_n \leq \left(\frac{(1-\lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1}\right)\|x_{n+1} - x_n\|^2 \ \forall n \geq 1. \tag{3. 33}$$

We claim that

$$\frac{(1-\lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1} \leq -\sigma \ \forall n \geq 1. \tag{3. 34}$$

Let be $n \geq 1$. Indeed, by the choice of ρ_n, it holds

$$\begin{aligned}
&\frac{(1-\lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1} \leq -\sigma \\
\iff &\lambda_n(\gamma_{n+1} + \sigma) + (\alpha_n\rho_n - 1)(1-\lambda_n) \leq 0 \\
\iff &\lambda_n(\gamma_{n+1} + \sigma) - \frac{\delta\lambda_n(1-\lambda_n)}{\alpha_n + \delta\lambda_n} \leq 0 \\
\iff &(\alpha_n + \delta\lambda_n)(\gamma_{n+1} + \sigma) + \delta\lambda_n \leq \delta.
\end{aligned}$$

By using (3. 32), we further get

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

where the last inequality follows by using the upper bound for $(\lambda_n)_{n\geq 1}$ in (3. 25). Hence, the claim in (3. 34) is true.

We obtain from (3. 33) and (3. 34) that

$$\mu_{n+1} - \mu_n \leq -\sigma\|x_{n+1} - x_n\|^2 \ \forall n \geq 1. \tag{3. 35}$$

Hence, the sequence $(\mu_n)_{n\geq 1}$ is nonincreasing and we take $M \geq 0$ an upper bound of it, that is $\mu_n \leq M$ for all $n \geq 1$. The bound for $(\alpha_n)_{n\geq 1}$ delivers

$$-\alpha\varphi_{n-1} \leq \varphi_n - \alpha\varphi_{n-1} \leq \mu_n \leq M \ \forall n \geq 1. \tag{3. 36}$$

We obtain

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

Combining (3. 35) and (3. 36), we get for all $n \geq 1$

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

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

(ii) We prove this statement by using the result of Opial given in Lemma 1.1. We have proven above that for an arbitrary $y \in \operatorname{Fix} T$ the inequality (3. 30) is true. By part (i), (3. 32) and Lemma 3.1 we derive that $\lim_{n\to+\infty}\|x_n - y\|$ exists (we take into consideration also that in (3. 30) $\alpha_n\rho_n < 1$ for all $n \geq 1$). On the other hand, let x be a weak sequential 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 (3. 26) we have

$$\begin{aligned} \|Tw_n - w_n\| &= \frac{1}{\lambda_n}\|x_{n+1} - w_n\| \\ &\leq \frac{1}{\lambda}\|x_{n+1} - w_n\| \\ &\leq \frac{1}{\lambda}\big(\|x_{n+1} - x_n\| + \alpha\|x_n - x_{n-1}\|\big), \end{aligned} \tag{3. 37}$$

thus by (i) we obtain $Tw_{n_k} - w_{n_k} \to 0$ as $k \to +\infty$. Applying now Lemma 1.3 for the sequence $(w_{n_k})_{k\in\mathbb{N}}$ we conclude that $x \in \operatorname{Fix} T$. Since the two assumptions of Lemma 1.1 are verified, it follows that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$. $\square$

Remark 3.5 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:

$$x_{n+1} = x_n + \lambda_n(Tx_n - x_n)\ \forall n \geq 1. \tag{3. 38}$$

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 [26, Theorem 5.14]).

Let us recall some technical results which are needed in the following.

If $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ are monotone, then we have the following characterization of the set of zeros of their sum (see [26, Proposition 25.1(ii)]):

$$\operatorname{zer}(A + B) = J_{\gamma B}(\operatorname{Fix} R_{\gamma A}R_{\gamma B})\ \forall \gamma > 0. \tag{3. 39}$$

The following result is a direct consequence of [26, Corollary 25.5] and it will be used in the proof of the convergence of the inertial Douglas–Rachford splitting algorithm.

Lemma 3.3 *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 are now in position to state the inertial Douglas–Rachford splitting algorithm and to present its convergence properties.

Theorem 3.5 *(Inertial Douglas–Rachford splitting algorithm) Let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone operators such that $\operatorname{zer}(A+B) \neq \varnothing$. 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 and fulfills

$$0 \leq \alpha_n \leq \alpha < 1\ \forall n \geq 1$$

and $\lambda, \sigma, \delta > 0$ *are such that*

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

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

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

(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* $\operatorname{zer}(A+B)$*.*

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

$$\begin{aligned}(\forall n \geq 1)\ x_{n+1} &= w_n + \lambda_n\Big[J_{\gamma A} \circ (2J_{\gamma B} - \operatorname{Id})w_n - J_{\gamma B}w_n\Big] \\ &= w_n + \lambda_n\left[\left(\frac{\operatorname{Id}+R_{\gamma A}}{2} \circ R_{\gamma B}\right)w_n - \frac{\operatorname{Id}+R_{\gamma B}}{2}w_n\right] \\ &= w_n + \frac{\lambda_n}{2}(Tw_n - w_n), \end{aligned} \quad (3.\ 40)$$

where $T := R_{\gamma A} \circ R_{\gamma B} : \mathcal{H} \to \mathcal{H}$ is a nonexpansive operator. From (3. 39) we have $\operatorname{zer}(A+B) = J_{\gamma B}(\operatorname{Fix} T)$, hence $\operatorname{Fix} T \neq \varnothing$. By applying Theorem 3.4, there exists $x \in \operatorname{Fix} T$ such that (i)-(iii) hold.

(iv) Follows from Theorem 3.4, (3. 37) and $z_n - y_n = \frac{1}{2}(Tw_n - w_n)$ 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

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

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

$$(z_n, u_n) \in \operatorname{gr}(\gamma A),\ (y_n, v_n) \in \operatorname{gr}(\gamma B) \text{ and } u_n + v_n = y_n - z_n \ \forall n \geq 1. \quad (3.\ 41)$$

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 3.3 we obtain $y \in \operatorname{zer}(\gamma A + \gamma B) = \operatorname{zer}(A+B)$, $(y, y-x) \in \operatorname{gr}\gamma A$ and $(y, x-y) \in \operatorname{gr}\gamma B$. 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 (3. 41) 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 (3. 41))

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

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 $n \to +\infty$. $\square$

Remark 3.6 In case $\alpha = 0$, which forces $\alpha_n = 0$ for all $n \geq 1$, the iterative scheme in Theorem 3.5 becomes the classical Douglas–Rachford splitting algorithm (see [26, Theorem 25.6]):

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

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 $\operatorname{zer}(A + B)$ has been for the first time reported in [127].

Remark 3.7 In case $Bx = 0$ for all $x \in \mathcal{H}$, the iterative scheme in Theorem 3.5 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 [4] as a proximal-point algorithm (see [122]) 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 [5, Theorem 2.1], [4] and [44, Remark 7].

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

We apply the inertial Douglas–Rachford algorithm proposed in the previous section 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 3.3 *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*

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in A\overline{x} + \sum_{i=1}^{m} L_i^*(B_i \Box D_i)(L_i\overline{x} - r_i) \tag{3. 42}$$

together with the dual inclusion

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,..., \overline{v}_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} z - \sum_{i=1}^m L_i^* \overline{v}_i \in Ax \\ \overline{v}_i \in (B_i \square D_i)(L_i x - r_i), \ i = 1,..., m. \end{cases} \tag{3. 43}$$

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 3.3, if

$$z - \sum_{i=1}^m L_i^* \overline{v}_i \in A\overline{x} \text{ and } \overline{v}_i \in (B_i \square D_i)(L_i \overline{x} - r_i), \ i = 1,..., m. \tag{3. 44}$$

Several particular instances of the primal-dual system of monotone inclusions (3. 42)–(3. 43) when applied to convex optimization problems can be found in [76, 130].

The inertial primal-dual Douglas-Rachford algorithm we would like to propose for solving (3. 42)–(3. 43) is formulated as follows.

Algorithm 3.1 *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 fulfilling $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ and the sequence $(\lambda_n)_{n\geq 1}$ be such that

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

For all $n \geq 1$ set

$$\left\lfloor \begin{array}{l} p_{1,n} = J_{\tau A}\left(x_n + \alpha_n(x_n - x_{n-1}) - \frac{\tau}{2}\sum_{i=1}^m L_i^*(v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1})) + \tau z\right) \\ w_{1,n} = 2p_{1,n} - x_n - \alpha_n(x_n - x_{n-1}) \\ \text{For } i = 1,..., m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = J_{\sigma_i B_i^{-1}}\left(v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1}) + \frac{\sigma_i}{2} L_i w_{1,n} - \sigma_i r_i\right) \\ w_{2,i,n} = 2p_{2,i,n} - v_{i,n} - \alpha_n(v_{i,n} - v_{i,n-1}) \end{array} \right. \\ z_{1,n} = w_{1,n} - \frac{\tau}{2}\sum_{i=1}^m L_i^* w_{2,i,n} \\ x_{n+1} = x_n + \alpha_n(x_n - x_{n-1}) + \lambda_n(z_{1,n} - p_{1,n}) \\ \text{For } i = 1,..., m \\ \quad \left\lfloor \begin{array}{l} z_{2,i,n} = J_{\sigma_i D_i^{-1}}\left(w_{2,i,n} + \frac{\sigma_i}{2} L_i(2z_{1,n} - w_{1,n})\right) \\ v_{i,n+1} = v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1}) + \lambda_n(z_{2,i,n} - p_{2,i,n}). \end{array} \right. \end{array} \right. \tag{3. 45}$$

Theorem 3.6 *In Problem 3.3, suppose that*

$$z \in \operatorname{ran}\left(A + \sum_{i=1}^m L_i^*(B_i \square D_i)(L_i \cdot - r_i)\right), \tag{3. 46}$$

and consider the sequences generated by Algorithm 3.1. Then there exists an element $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 ... \times \mathcal{G}_m$ 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 3.3;*

(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 the sequences* $(p_{1,n}, p_{2,1,n},..., p_{2,m,n})_{n\geq 1}$ *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 3.3.*

Proof. For the proof we use Theorem 3.5 and adapt the techniques from [59] (see also [130]) to the given settings. We consider the Hilbert space $\boldsymbol{\mathcal{K}} = \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ endowed with inner product and associated norm defined, for $(x, v_1,..., v_m)$, $(y, q_1,..., q_m) \in \boldsymbol{\mathcal{K}}$, via

$$\begin{gathered} \langle (x, v_1,..., v_m), (y, q_1,..., q_m)\rangle_{\boldsymbol{\mathcal{K}}} = \langle x, y\rangle_{\mathcal{H}} + \sum_{i=1}^{m} \langle v_i, q_i\rangle_{\mathcal{G}_i} \\ \text{and} \\ \|(x, v_1,..., v_m)\|_{\boldsymbol{\mathcal{K}}} = \sqrt{\|x\|_{\mathcal{H}}^2 + \sum_{i=1}^{m} \|v_i\|_{\mathcal{G}_i}^2}, \end{gathered} \tag{3. 47}$$

respectively. Furthermore, we consider the set-valued operator

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

which is maximally monotone, since A and B_i, $i = 1,..., m$, are maximally monotone (see [26, Proposition 20.22 and Proposition 20.23]), and the linear continuous operator

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

which is skew-symmetric (i. e. $\boldsymbol{S}^* = -\boldsymbol{S}$) and hence maximally monotone (see [26, Example 20.30]). Moreover, we consider the set-valued operator

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

which is once again maximally monotone, since D_i is maximally monotone for $i = 1,..., m$. Therefore, since $\operatorname{dom} \boldsymbol{S} = \boldsymbol{\mathcal{K}}$, both $\frac{1}{2}\boldsymbol{S} + \boldsymbol{Q}$ and $\frac{1}{2}\boldsymbol{S} + \boldsymbol{M}$ are maximally monotone (see [26, Corollary 24.4(i)]). Furthermore, one can easily notice that

$$(3.\ 46) \Leftrightarrow \operatorname{zer}(\boldsymbol{M} + \boldsymbol{S} + \boldsymbol{Q}) \neq \varnothing$$

and

$$\begin{aligned} &(x, v_1,..., v_m) \in \operatorname{zer}(\boldsymbol{M} + \boldsymbol{S} + \boldsymbol{Q}) \\ \Rightarrow &(x, v_1,..., v_m) \text{ is a primal-dual solution to Problem 3.3.} \end{aligned} \tag{3. 48}$$

We also introduce the linear continuous operator

$$\boldsymbol{V}:\boldsymbol{\mathcal{K}}\to\boldsymbol{\mathcal{K}},\quad (x,v_1,\ldots,v_m)\mapsto\left(\frac{x}{\tau}-\frac{1}{2}\sum_{i=1}^m L_i^*v_i,\frac{v_1}{\sigma_1}-\frac{1}{2}L_1x,\ldots,\frac{v_m}{\sigma_m}-\frac{1}{2}L_mx\right),$$

which is self-adjoint and ρ-strongly positive (see [59]) for

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

namely, the following inequality holds

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

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

Note that for all $n\geq 1$, the algorithmic scheme (3. 45) is equivalent to

$$\left\lfloor\begin{array}{l}
\frac{x_n-p_{1,n}}{\tau}+\alpha_n\frac{x_n-x_{n-1}}{\tau}-\frac{1}{2}\sum_{i=1}^m L_i^*(v_{i,n}+\alpha_n(v_{i,n}-v_{i,n-1}))\in Ap_{1,n}-z\\
w_{1,n}=2p_{1,n}-x_n-\alpha_n(x_n-x_{n-1})\\
\text{For } i=1,\ldots,m\\
\quad\left\lfloor\begin{array}{l}
\frac{v_{i,n}-p_{2,i,n}}{\sigma_i}+\alpha_n\frac{v_{i,n}-v_{i,n-1}}{\sigma_i}-\frac{1}{2}L_i(x_n-p_{1,n}+\alpha_n(x_n-x_{n-1}))\\
\qquad\qquad\qquad\qquad\qquad\qquad\in-\frac{1}{2}L_ip_{1,n}+B_i^{-1}p_{2,i,n}+r_i\\
w_{2,i,n}=2p_{2,i,n}-v_{i,n}-\alpha_n(v_{i,n}-v_{i,n-1})
\end{array}\right.\\
\frac{w_{1,n}-z_{1,n}}{\tau}-\frac{1}{2}\sum_{i=1}^m L_i^*w_{2,i,n}=0\\
x_{n+1}=x_n+\alpha_n(x_n-x_{n-1})+\lambda_n(z_{1,n}-p_{1,n})\\
\text{For } i=1,\ldots,m\\
\quad\left\lfloor\begin{array}{l}
\frac{w_{2,i,n}-z_{2,i,n}}{\sigma_i}-\frac{1}{2}L_i(w_{1,n}-z_{1,n})\in-\frac{1}{2}L_iz_{1,n}+D_i^{-1}z_{2,i,n}\\
v_{i,n+1}=v_{i,n}+\alpha_n(v_{i,n}-v_{i,n-1})+\lambda_n(z_{2,i,n}-p_{2,i,n}).
\end{array}\right.
\end{array}\right. \tag{3. 49}$$

By considering for all $n\geq 1$ the notations

$$\boldsymbol{x}_n=(x_n,v_{1,n},\ldots,v_{m,n}),$$

$$\boldsymbol{y}_n=(p_{1,n},p_{2,1,n},\ldots,p_{2,m,n})$$

and

$$\boldsymbol{z}_n=(z_{1,n},z_{2,1,n},\ldots,z_{2,m,n}),$$

the scheme (3. 49) can equivalently be written in the form

$$(\forall n\geq 1)\left\lfloor\begin{array}{l}
\boldsymbol{V}(\boldsymbol{x}_n-\boldsymbol{y}_n+\alpha_n(\boldsymbol{x}_n-\boldsymbol{x}_{n-1}))\in\left(\frac{1}{2}\boldsymbol{S}+\boldsymbol{M}\right)\boldsymbol{y}_n\\
\boldsymbol{V}(2\boldsymbol{y}_n-\boldsymbol{x}_n-\boldsymbol{z}_n-\alpha_n(\boldsymbol{x}_n-\boldsymbol{x}_{n-1}))\in\left(\frac{1}{2}\boldsymbol{S}+\boldsymbol{Q}\right)\boldsymbol{z}_n\\
\boldsymbol{x}_{n+1}=\boldsymbol{x}_n+\alpha_n(\boldsymbol{x}_n-\boldsymbol{x}_{n-1})+\lambda_n\left(\boldsymbol{z}_n-\boldsymbol{y}_n\right),
\end{array}\right. \tag{3. 50}$$

which is equivalent to

$$(\forall n\geq 1)\left\lfloor\begin{array}{l}
\boldsymbol{y}_n=\left(\mathrm{Id}+\boldsymbol{V}^{-1}(\frac{1}{2}\boldsymbol{S}+\boldsymbol{M})\right)^{-1}(\boldsymbol{x}_n+\alpha_n(\boldsymbol{x}_n-\boldsymbol{x}_{n-1}))\\
\boldsymbol{z}_n=\left(\mathrm{Id}+\boldsymbol{V}^{-1}(\frac{1}{2}\boldsymbol{S}+\boldsymbol{Q})\right)^{-1}(2\boldsymbol{y}_n-\boldsymbol{x}_n-\alpha_n(\boldsymbol{x}_n-\boldsymbol{x}_{n-1}))\\
\boldsymbol{x}_{n+1}=\boldsymbol{x}_n+\alpha_n(\boldsymbol{x}_n-\boldsymbol{x}_{n-1})+\lambda_n\left(\boldsymbol{z}_n-\boldsymbol{y}_n\right),
\end{array}\right. \tag{3. 51}$$

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

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle_{\boldsymbol{\mathcal{K}}_{\boldsymbol{V}}}=\langle\boldsymbol{x},\boldsymbol{V}\boldsymbol{y}\rangle_{\boldsymbol{\mathcal{K}}}\text{ and }\|\boldsymbol{x}\|_{\boldsymbol{\mathcal{K}}_{\boldsymbol{V}}}=\sqrt{\langle\boldsymbol{x},\boldsymbol{V}\boldsymbol{x}\rangle_{\boldsymbol{\mathcal{K}}}}. \tag{3. 52}$$

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

$$\boldsymbol{B} := \boldsymbol{V}^{-1}\left(\frac{1}{2}\boldsymbol{S}+\boldsymbol{M}\right) \text{ and } \boldsymbol{A} := \boldsymbol{V}^{-1}\left(\frac{1}{2}\boldsymbol{S}+\boldsymbol{Q}\right) \tag{3. 53}$$

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

Taking this into account, it shows that (3. 51) becomes

$$(\forall n \geq 1) \left\lfloor \begin{array}{l} \boldsymbol{y}_n = J_{\boldsymbol{B}}\left(\boldsymbol{x}_n + \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})\right) \\ \boldsymbol{z}_n = J_{\boldsymbol{A}}\left(2\boldsymbol{y}_n - \boldsymbol{x}_n - \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1})\right) \\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n(\boldsymbol{x}_n - \boldsymbol{x}_{n-1}) + \lambda_n\left(\boldsymbol{z}_n - \boldsymbol{y}_n\right), \end{array} \right. \tag{3. 54}$$

which is the inertial Douglas–Rachford algorithm presented in Theorem 3.5 in the space $\boldsymbol{\mathcal{K}}_{\boldsymbol{V}}$ for $\gamma = 1$. Furthermore, we have

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

(i) By Theorem 3.5 (i), there exists $\overline{\boldsymbol{x}} = (\overline{x}, \overline{v}_1, ..., \overline{v}_m) \in \operatorname{Fix}(R_{\boldsymbol{A}}R_{\boldsymbol{B}})$, such that $J_{\boldsymbol{B}}\overline{\boldsymbol{x}} \in \operatorname{zer}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{zer}(\boldsymbol{M}+\boldsymbol{S}+\boldsymbol{Q})$. The claim follows from (3. 48) and by identifying $J_{\boldsymbol{B}}\overline{\boldsymbol{x}}$.

(ii) Since $\boldsymbol{V}$ is ρ-strongly positive, we obtain from Theorem 3.5 (ii) that

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

and therefore the claim follows by considering (3. 47).

(iii)–(vi) Follows directly from Theorem 3.5 (iii)–(vi).

(vii) The uniform monotonicity of A and B_i^{-1}, $i = 1,...,m$, implies uniform monotonicity of $\boldsymbol{M}$ on $\boldsymbol{\mathcal{K}}$ (see, for instance, [59, Theorem 2.1 (ii)]), while this further implies uniform monotonicity of $\boldsymbol{B}$ on $\boldsymbol{\mathcal{K}}_{\boldsymbol{V}}$. Therefore, the claim follows from Theorem 3.5 (vii). □

3.2.3 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 deal with the following problem.

Problem 3.4 *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*

$$(P) \quad \inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^m (g_i \Box l_i)(L_i x - r_i) - \langle x, z\rangle \right\} \tag{3. 55}$$

and its conjugate dual problem

$$(D) \quad \sup_{(v_1,...,v_m)\in\mathcal{G}_1\times...\times\mathcal{G}_m} \left\{ -f^*\left(z - \sum_{i=1}^m L_i^* v_i\right) - \sum_{i=1}^m \left(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i\rangle\right) \right\}. \tag{3. 56}$$

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 (3. 42) reads

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } z \in \partial f(\overline{x}) + \sum_{i=1}^{m} L_i^*(\partial g_i \Box \partial l_i)(L_i\overline{x} - r_i), \tag{3. 57}$$

while the dual inclusion problem (3. 43) reads

$$\text{find } \overline{v}_1 \in \mathcal{G}_1,\ ..., \overline{v}_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in \partial f(x) \\ \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i x - r_i),\ i = 1,..., m. \end{cases} \tag{3. 58}$$

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 (3. 57)–(3. 58), namely,

$$z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in \partial f(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i\overline{x} - r_i),\ i = 1,..., m, \tag{3. 59}$$

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

Combining this statement with Algorithm 3.1 and Theorem 3.6 gives rise to the following iterative scheme and corresponding convergence theorem for the primal-dual pair of optimization problems (P)–(D).

Algorithm 3.2 *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 fulfilling $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ and the sequence $(\lambda_n)_{n\geq 1}$ be such that

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

For all $n \geq 1$ set

$$\left\lfloor \begin{array}{l} p_{1,n} = \operatorname{prox}_{\tau f}\big(x_n + \alpha_n(x_n - x_{n-1}) - \frac{\tau}{2}\sum_{i=1}^{m} L_i^*(v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1})) + \tau z\big) \\ w_{1,n} = 2p_{1,n} - x_n - \alpha_n(x_n - x_{n-1}) \\ \textit{For } i = 1,..., m \\ \quad \left\lfloor \begin{array}{l} p_{2,i,n} = \operatorname{prox}_{\sigma_i g_i^*}\big(v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1}) + \frac{\sigma_i}{2}L_i w_{1,n} - \sigma_i r_i\big) \\ w_{2,i,n} = 2p_{2,i,n} - v_{i,n} - \alpha_n(v_{i,n} - v_{i,n-1}) \end{array} \right. \\ z_{1,n} = w_{1,n} - \frac{\tau}{2}\sum_{i=1}^{m} L_i^* w_{2,i,n} \\ x_{n+1} = x_n + \alpha_n(x_n - x_{n-1}) + \lambda_n(z_{1,n} - p_{1,n}) \\ \textit{For } i = 1,..., m \\ \quad \left\lfloor \begin{array}{l} z_{2,i,n} = \operatorname{prox}_{\sigma_i l_i^*}\big(w_{2,i,n} + \frac{\sigma_i}{2}L_i(2z_{1,n} - w_{1,n})\big) \\ v_{i,n+1} = v_{i,n} + \alpha_n(v_{i,n} - v_{i,n-1}) + \lambda_n(z_{2,i,n} - p_{2,i,n}). \end{array} \right. \end{array} \right. \tag{3. 60}$$

Theorem 3.7 *In Problem 3.4, suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^{m} L_i^*(\partial g_i \Box \partial l_i)(L_i \cdot - r_i)\right), \tag{3. 61}$$

and consider the sequences generated by Algorithm 3.2. Then there exists an element $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 ... \times \mathcal{G}_m$ such that the following statements are true:

(i) *By setting*

$$\overline{p}_1 = \operatorname{prox}_{\tau f}\left(\overline{x} - \frac{\tau}{2}\sum_{i=1}^{m} L_i^* \overline{v}_i + \tau z\right),$$
$$\overline{p}_{2,i} = \operatorname{prox}_{\sigma_i g_i^*}\left(\overline{v}_i + \frac{\sigma_i}{2} L_i(2\overline{p}_1 - \overline{x}) - \sigma_i r_i\right),\ i = 1,..., m,$$

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 3.4, hence* $\overline{p}_1$ *is an optimal solution to* (P) *and* $(\overline{p}_{2,1},..., \overline{p}_{2,m})$ *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) $(p_{1,n}, p_{2,1,n},..., p_{2,m,n})_{n\geq 1}$ *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* f *and* g_i^*, $i = 1,..., m$, *are uniformly convex, 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 3.4.*

Remark 3.8 Considering the setting of Problem 3.4, the hypothesis (3. 61) in the above theorem is fulfilled if the primal problem (3. 55) has an optimal solution, the regularity condition (2. 42) holds and

$$0 \in \operatorname{sqri}(\operatorname{dom} g_i^* - \operatorname{dom} l_i^*) \text{ for } i = 1, ..., m.$$

According to [26, Proposition 15.7], the latter guarantees that $\Gamma(\mathcal{G}_i)$, $i = 1, ..., m$.

3.2.4 Numerical experiments

Clustering

We consider again a numerical experiment in cluster analysis, where one can observe a better performance of the inertial Douglas-Rachford algorithm in comparison with the noninertial one. We briefly recall some notations used in clustering and we refer to Chapter 2 for other details concerning this application. 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 [73, 96], clustering can be formulated as the convex optimization problem

$$\inf_{x_i\in\mathbb{R}^n,\, i=1,\ldots,m} \left\{ \frac{1}{2}\sum_{i=1}^{m} \|x_i - u_i\|^2 + \gamma \sum_{i<j} \omega_{ij}\|x_i - x_j\|_p \right\}, \tag{3. 62}$$

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 (2. 47).

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 (2. 47) can be equivalently written as

$$\inf_{x\in\mathbb{R}^{mn}} \{h(x) + g(Ax)\}, \tag{3. 63}$$

	$p = 2$, $\gamma = 5.2$		$p = 1$, $\gamma = 4$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$
Algorithm 3.2	0.65s (175)	1.36s (371)	0.63s (176)	1.27s (374)
DR [59]	0.78s (216)	1.68s (460)	0.78s (218)	1.68s (464)
FB [130]	2.48s (1353)	5.72s (3090)	2.01s (1092)	4.05s (2226)
FB Acc [53]	2.04s (1102)	4.11s (2205)	1.74s (950)	3.84s (2005)
FBF [76]	7.67s (2123)	17.58s (4879)	6.33s (1781)	13.22s (3716)
FBF Acc [58]	5.05s (1384)	10.27s (2801)	4.83s (1334)	9.98s (2765)
PD [69]	1.48s (780)	3.26s (1708)	1.44s (772)	3.18s (1722)
PD Acc [69]	1.28s (671)	3.14s (1649)	1.23s (665)	3.12s (1641)
Nesterov [107]	7.85s (3811)	42.69s (21805)	7.46s (3936)	> 190s ($> 10^5$)
FISTA [28]	7.55s (4055)	51.01s (27356)	6.55s (3550)	47.81s (26069)

Table 3.1: 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.

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 3.1). 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 3.1, it shows that Algorithm 3.2 performs better than the noninertial Douglas–Rachford (DR) method proposed in [59, 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 [76], and the forward-backward (FB) method from [130], where in both methods the function h is processed via a forward step. The accelerated versions of the latter and of the primal-dual (PD) method from [69] converge in less time than their regular variants, but are still slower than Algorithm 3.1. Notice that the methods called Nesterov and FISTA are accelerated proximal gradient algorithms which are applied to the Fenchel dual problem to (3. 63).

The generalized Heron problem

In the sequel we investigate the *generalized Heron problem* which has been recently investigated in [102, 103] and where for its solving subgradient-type methods have

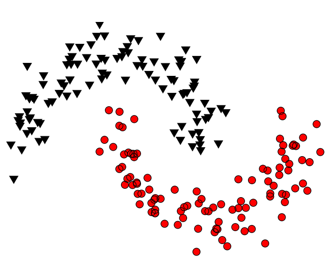


Figure 3.1: Clustering two interlocking half moons. The colors (resp. the shapes) show the correct affiliations.

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

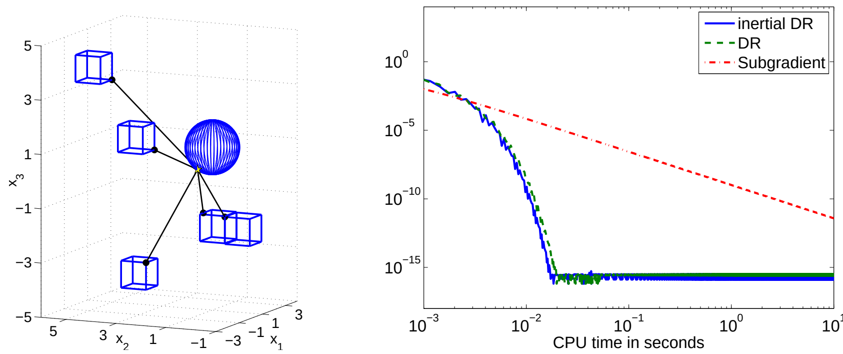


Figure 3.2: Generalized Heron problem with cubes and ball constraint set on the left-hand side, performance evaluation for the RMSE on the right-hand side.

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

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

Thus the *generalized Heron problem* reads

$$\inf_{x\in\Omega} \sum_{i=1}^{m} d(x;\Omega_i), \tag{3. 64}$$

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 (3. 64) perfectly fits into the framework considered in Problem 3.4 when setting

$$f = \delta_\Omega, \text{ and } g_i = \|\cdot\|, \ l_i = \delta_{\Omega_i} \text{ for all } i = 1, \ldots, m. \tag{3. 65}$$

However, note that (3. 64) cannot be solved via the primal-dual methods in [76] and [130], 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

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

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 (1. 33).

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.

	Algorithm 3.2		Douglas–Rachford, [59]		Subgradient, [102, 103]	
	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-10}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-10}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-10}$
$n = 2,\ m = 5$	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 3.2: 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.

Table 3.2 shows a comparison between Algorithm 3.2, the Douglas–Rachford type method from [59, Algorithm 3.1], and the subgradient approach described in [102, 103] when applied to different instances of the generalized Heron problem. One such particular case is displayed in Figure 3.2 when $n = 3$ and $m = 5$, while the evolution of the RMSE values is given there in more detail. Empty cells in Table 3.2 indicate that it took more than 60 seconds to pass the stopping criterion. Based on the provided data, one can say that both Algorithm 3.2 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.

3.3 Splitting algorithms for nonconvex optimization problems

The extension of proximal-type algorithms and of the corresponding convergence analysis to the nonconvex setting is a challenging ongoing research topic. By assuming that the functions in the objective share some analytic features and by making consequently use of a generalization to the nonsmooth setting of the Kurdyka-Łojasiewicz property initially introduced for smooth functions, the proximal point algorithm for minimizing a proper and lower semicontinuous function and the forward-backward scheme for minimizing the sum of a nonsmooth lower semicontinuous function with a smooth one have proved to possess good convergence properties also in the nonconvex case, see Attouch and Bolte [9], Attouch, Bolte, Redont and Soubeyran [10], Attouch, Bolte and Svaiter [11], Bolte, Sabach and Teboulle [35], Chouzenoux, Pesquet and Repetti [74], Frankel, Farrigos and Peypouquet [88] (wee mention here also the work of Noll [108] concerning descent methods). The particular class of functions fulfilling the Kurdyka-Łojasiewicz property includes semi-algebraic functions, real subanalytic functions, semi-convex functions, uniformly convex functions, etc. (see also [34, 94, 98]).

The interest of having convergence properties in the nonconvex setting is motivated among others by applications in connection to sparse nonnegative matrix factorization, hard constrained feasibility, compressive sensing, etc. In what regards the latter, they give rise to the solving of optimization problems where the sparsity measure is used as regularization functional. Due to the fact that this functional is semi-algebraic, algorithms for solving nonsmooth optimization problems involving KL functions represent a serious option in this sense (see [11, Example 5.4]).

Throughout this section, we consider on $\mathbb{R}^m$ (where $m \geq 1$) the Euclidean scalar product and the induced norm denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. Notice that all the finite-dimensional spaces considered below are endowed with the topology induced by the Euclidean norm.

For the following generalized subdifferential notions and their basic properties we refer to [101, 123]. Let $f:\mathbb{R}^m\rightarrow(-\infty,+\infty]$ be a proper and lower semicontinuous function. If $x\in\operatorname{dom}f$, we consider the *Fréchet (viscosity) subdifferential* of f at x as being the set

$$\hat{\partial}f(x)=\left\{v\in\mathbb{R}^m:\liminf_{y\rightarrow x}\frac{f(y)-f(x)-\langle v,y-x\rangle}{\|y-x\|}\geq 0\right\}.$$

For $x\notin\operatorname{dom}f$ we set $\hat{\partial}f(x):=\emptyset$. The *limiting (Mordukhovich) subdifferential* is defined at $x\in\operatorname{dom}f$ by

$$\partial f(x)=\{v\in\mathbb{R}^m:\exists x_n\rightarrow x,f(x_n)\rightarrow f(x)\text{ and }\exists v_n\in\hat{\partial}f(x_n),v_n\rightarrow v\text{ as }n\rightarrow+\infty\},$$

while for $x\notin\operatorname{dom}f$, one takes $\partial f(x):=\emptyset$.

Notice that in case f is convex, these notions coincide with the *convex subdifferential*, which means that $\hat{\partial}f(x)=\partial f(x)=\{v\in\mathbb{R}^m:f(y)\geq f(x)+\langle v,y-x\rangle\ \forall y\in\mathbb{R}^m\}$ for all $x\in\operatorname{dom}f$.

It holds $\hat{\partial}f(x)\subseteq\partial f(x)$ for each $x\in\mathbb{R}^m$. We will use the following closedness criteria concerning the graph of the limiting subdifferential: if $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{R}^m$ such that $v_n\in\partial f(x_n)$ for all $n\in\mathbb{N}$, $(x_n,v_n)\rightarrow(x,v)$ and $f(x_n)\rightarrow f(x)$ as $n\rightarrow+\infty$, then $v\in\partial f(x)$.

The Fermat rule reads in this nonsmooth setting as: if $x\in\mathbb{R}^m$ is a local minimizer of f, then $0\in\partial f(x)$. Notice that in case f is continuously differentiable around $x\in\mathbb{R}^m$ we have $\partial f(x)=\{\nabla f(x)\}$. Let us denote by

$$\operatorname{crit}(f)=\{x\in\mathbb{R}^m:0\in\partial f(x)\}$$

the set of *(limiting)-critical points* of f. We mention also the following subdifferential rule: if $f:\mathbb{R}^m\rightarrow(-\infty,+\infty]$ is proper and lower semicontinuous and $h:\mathbb{R}^m\rightarrow\mathbb{R}$ is a continuously differentiable function, then $\partial(f+h)(x)=\partial f(x)+\nabla h(x)$ for all $x\in\mathbb{R}^m$.

We turn now our attention to the class of functions satisfying the *Kurdyka-Łojasiewicz property.* This class of functions will play a crucial role when proving the convergence of the proposed inertial algorithm. For $\eta\in(0,+\infty]$, we denote by Θ_η the class of concave and continuous functions $\varphi:[0,\eta)\rightarrow[0,+\infty)$ such that $\varphi(0)=0$, φ is continuously differentiable on $(0,\eta)$, continuous at 0 and $\varphi'(s)>0$ for all $s\in(0,\eta)$. In the following definition (see [10, 35]) we use also the *distance function* to a set, defined for $A\subseteq\mathbb{R}^m$ as $\operatorname{dist}(x,A)=\inf_{y\in A}\|x-y\|$ for all $x\in\mathbb{R}^m$.

Definition 3.1 (*Kurdyka-Łojasiewicz property*) Let $f:\mathbb{R}^m\rightarrow(-\infty,+\infty]$ be a proper and lower semicontinuous function. We say that f satisfies the *Kurdyka-Łojasiewicz (KL) property* at $\overline{x}\in\operatorname{dom}\partial f=\{x\in\mathbb{R}^m:\partial f(x)\neq\emptyset\}$ if there exists $\eta\in(0,+\infty]$, a neighborhood U of $\overline{x}$ and a function $\varphi\in\Theta_\eta$ such that for all x in the intersection

$$U\cap\{x\in\mathbb{R}^m:f(\overline{x})<f(x)<f(\overline{x})+\eta\}$$

the following inequality holds

$$\varphi'(f(x)-f(\overline{x}))\operatorname{dist}(0,\partial f(x))\geq 1.$$

If f satisfies the KL property at each point in $\operatorname{dom}\partial f$, then f is called a *KL function.*

The origins of this notion go back to the pioneering work of Łojasiewicz [98], where it is proved that for a real-analytic function $f : \mathbb{R}^m \to \mathbb{R}$ and a critical point $\overline{x} \in \mathbb{R}^m$ (that is $\nabla f(\overline{x}) = 0$), there exists $\theta \in [1/2, 1)$ such that the function $|f - f(\overline{x})|^\theta \|\nabla f\|^{-1}$ is bounded around $\overline{x}$. This corresponds to the situation when $\varphi(s) = Cs^{1-\theta}$, where $C > 0$. The result of Łojasiewicz allows the interpretation of the KL property as a re-parametrization of the function values in order to avoid flatness around the critical points. Kurdyka [94] extended this property to differentiable functions definable in an o-minimal structure. Further extensions to the nonsmooth setting can be found in [10, 32–34].

One of the remarkable properties of the KL functions is their ubiquity in applications (see [35]). To the class of KL functions belong semi-algebraic, real subanalytic, semiconvex, uniformly convex and convex functions satisfying a growth condition. We refer the reader to [9–11, 32–35] and the references therein for more details regarding all the classes mentioned above and illustrating examples.

An important role in our convergence analysis will be played by the following uniformized KL property given in [35, Lemma 6].

Lemma 3.4 *Let $\Omega \subseteq \mathbb{R}^m$ be a compact set and let $f : \mathbb{R}^m \to (-\infty, +\infty]$ be a proper and lower semicontinuous function. Assume that f is constant on Ω and f satisfies the KL property at each point of Ω. Then there exist $\varepsilon, \eta > 0$ and $\varphi \in \Theta_\eta$ such that for all $\overline{x} \in \Omega$ and for all x in the intersection*

$$\{x \in \mathbb{R}^m : \operatorname{dist}(x, \Omega) < \varepsilon\} \cap \{x \in \mathbb{R}^m : f(\overline{x}) < f(x) < f(\overline{x}) + \eta\} \tag{3. 66}$$

the following inequality holds

$$\varphi'(f(x) - f(\overline{x})) \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{3. 67}$$

Finally, let us present a convergence result (see for example [47]) which will be used in the convergence analysis below.

Lemma 3.5 *Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be nonnegative real sequences, such that $\sum_{n\in\mathbb{N}} b_n < +\infty$ and $a_{n+1} \leq a \cdot a_n + b \cdot a_{n-1} + b_n$ for all $n \geq 1$, where $a \in \mathbb{R}$, $b \geq 0$ and $a + b < 1$. Then $\sum_{n\in\mathbb{N}} a_n < +\infty$.*

3.3.1 An inertial forward-backward algorithm in the nonconvex setting

In this section we present an inertial forward-backward algorithm for solving a fully nonconvex optimization problem and study its convergence properties. The problem under investigation has the following formulation.

Problem 3.5 *Let $f : \mathbb{R}^m \to (-\infty, +\infty]$ be a proper, lower semicontinuous function which is bounded from below and let $g : \mathbb{R}^m \to \mathbb{R}$ be a Fréchet differentiable function with $L_{\nabla g}$-Lipschitz continuous gradient, where $L_{\nabla g} \geq 0$. We deal with the optimization problem*

$$(P) \quad \inf_{x\in\mathbb{R}^m} [f(x) + g(x)]. \tag{3. 68}$$

In the iterative scheme we propose below, we use also the function $F : \mathbb{R}^m \to \mathbb{R}$, assumed to be $\sigma-$strongly convex, Fréchet differentiable and such that ∇F is $L_{\nabla F}$-Lipschitz continuous, where $\sigma, L_{\nabla F} > 0$. The *Bregman distance* to F, denoted by $D_F : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, is defined as

$$D_F(x, y) = F(x) - F(y) - \langle \nabla F(y), x - y \rangle \ \forall (x, y) \in \mathbb{R}^m \times \mathbb{R}^m.$$

Notice that the properties of the function F ensure the following inequalities

$$\frac{\sigma}{2}\|x-y\|^2 \le D_F(x,y) \le \frac{L_{\nabla F}}{2}\|x-y\|^2 \ \forall x,y \in \mathbb{R}^m. \tag{3. 69}$$

We propose the following iterative scheme for solving (3. 68).

Algorithm 3.3 *Chose $x_0, x_1 \in \mathbb{R}^m$, $\underline{\alpha}, \overline{\alpha} > 0$, $\beta \ge 0$ and sequences $(\alpha_n)_{n\ge 1}, (\beta_n)_{n\ge 1}$ fulfilling*

$$0 < \underline{\alpha} \le \alpha_n \le \overline{\alpha} \ \forall n \ge 1$$

and

$$0 \le \beta_n \le \beta \ \forall n \ge 1.$$

For all $n \ge 1$, we consider the iterative scheme

$$x_{n+1} \in \operatorname*{argmin}_{u\in\mathbb{R}^m} \left\{ D_F(u,x_n) + \alpha_n\langle u, \nabla g(x_n)\rangle + \beta_n\langle u, x_{n-1}-x_n\rangle + \alpha_n f(u) \right\}. \tag{3. 70}$$

Due to the subdifferential sum formula mentioned in the previous section, one can see that any sequence generated by this algorithm satisfies the relation

$$x_{n+1} \in (\nabla F + \alpha_n \partial f)^{-1}(\nabla F(x_n) - \alpha_n \nabla g(x_n) + \beta_n(x_n - x_{n-1})) \ \forall n \ge 1. \tag{3. 71}$$

Further, since f is proper, lower semicontinuous and bounded from below and D_F is coercive in its first argument (that is $\lim_{\|x\|\to+\infty} D_F(x,y) = +\infty$ for all $y \in \mathbb{R}^m$), the iterative scheme is well-defined, meaning that the existence of x_n is guaranteed for each $n \ge 2$, since the objective function in the minimization problem to be solved at each iteration is coercive.

Remark 3.9 The assumption that f should be bounded from below is imposed in order to ensure that in each iteration one can chose at least one x_n (that is the argmin in (3. 70) is nonempty). One can replace this requirement by asking that the objective function in the minimization problem considered in (3. 70) is coercive and the theory presented below still remains valid. This observation is useful when dealing with optimization problems as the ones considered in Subsection 3.3.2.

Before proceeding with the convergence analysis, we discuss the relation of our scheme to other algorithms from the literature. Let us take first $F(x) = \frac{1}{2}\|x\|^2$ for all $x \in \mathbb{R}^m$. In this case $D_F(x,y) = \frac{1}{2}\|x-y\|^2$ for all $(x,y) \in \mathbb{R}^m \times \mathbb{R}^m$ and $\sigma = L_{\nabla F} = 1$. The iterative scheme becomes

$$(\forall n \ge 1) \ \ x_{n+1} \in \operatorname*{argmin}_{u\in\mathbb{R}^m} \left\{ \frac{\|u-(x_n - \alpha_n\nabla g(x_n) + \beta_n(x_n - x_{n-1}))\|^2}{2\alpha_n} + f(u) \right\}. \tag{3. 72}$$

A similar inertial type algorithm has been analyzed in [110], however in the restrictive case when f is convex. If we take in addition $\beta = 0$, which enforces $\beta_n = 0$ for all $n \ge 1$, then (3. 72) becomes

$$(\forall n \ge 1) \ \ x_{n+1} \in \operatorname*{argmin}_{u\in\mathbb{R}^m} \left\{ \frac{\|u-(x_n - \alpha_n\nabla g(x_n))\|^2}{2\alpha_n} + f(u) \right\}, \tag{3. 73}$$

the convergence of which has been investigated in [35] in the full nonconvex setting. Notice that forward-backward algorithms with variable metrics for KL functions have been proposed in [74, 88].

On the other hand, if we take $g(x) = 0$ for all $x \in \mathbb{R}^m$, the iterative scheme in (3. 72) becomes

$$(\forall n \geq 1) \;\; x_{n+1} \in \underset{u \in \mathbb{R}^m}{\operatorname{argmin}} \left\{ \frac{\|u - (x_n + \beta_n(x_n - x_{n-1}))\|^2}{2\alpha_n} + f(u) \right\}, \tag{3. 74}$$

which is a proximal point algorithm with inertial/memory effects formulated in the nonconvex setting designed for finding the critical points of f. The iterative scheme without the inertial term, that is when $\beta = 0$ and, so, $\beta_n = 0$ for all $n \geq 1$, has been considered in the context of KL functions in [9].

Let us mention that in the full convex setting, which means that f and g are convex functions, in which case for all $n \geq 2$, x_n is uniquely determined and can be expressed via the *proximal operator* of f, (3. 72) can be derived from the iterative scheme proposed in [104], (3. 73) is the classical forward-backward algorithm (see for example [26] or [75]) and (3. 74) has been analyzed in [5] in the more general context of monotone inclusion problems.

Let us start now with the investigation of the convergence of the proposed algorithm.

Lemma 3.6 *In the setting of Problem 3.5, let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.3. Then one has*

$$(f+g)(x_{n+1}) + M_1\|x_n - x_{n+1}\|^2 \leq (f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2 \;\; \forall n \geq 1,$$

where

$$M_1 = \frac{\sigma - \overline{\alpha} L_{\nabla g}}{2\overline{\alpha}} - \frac{\beta}{2\underline{\alpha}} \text{ and } M_2 = \frac{\beta}{2\underline{\alpha}}. \tag{3. 75}$$

Moreover, for $0 < \underline{\alpha} \leq \overline{\alpha}$ and $\beta > 0$ satisfying

$$\sigma > \overline{\alpha} L_{\nabla g} + 2\beta \frac{\overline{\alpha}}{\underline{\alpha}}, \tag{3. 76}$$

one has $M_1 > M_2$.

Proof. Let be $n \geq 1$ fixed. Due to (3. 70) we have

$$\begin{aligned} &D_F(x_{n+1}, x_n) + \alpha_n \langle x_{n+1}, \nabla g(x_n)\rangle + \beta_n \langle x_{n+1}, x_{n-1} - x_n\rangle + \alpha_n f(x_{n+1}) \\ \leq\;& D_F(x_n, x_n) + \alpha_n \langle x_n, \nabla g(x_n)\rangle + \beta_n \langle x_n, x_{n-1} - x_n\rangle + \alpha_n f(x_n) \end{aligned}$$

or, equivalently,

$$\begin{aligned} D_F(x_{n+1}, x_n) + \langle x_{n+1} - x_n, \alpha_n \nabla g(x_n) - \beta_n(x_n - x_{n-1})\rangle + \alpha_n f(x_{n+1}) \\ \leq \alpha_n f(x_n). \end{aligned} \tag{3. 77}$$

On the other hand, by Lemma 1.4 we have

$$\langle \nabla g(x_n), x_{n+1} - x_n\rangle \geq g(x_{n+1}) - g(x_n) - \frac{L_{\nabla g}}{2}\|x_n - x_{n+1}\|^2.$$

At the same time

$$\langle x_{n+1} - x_n, x_{n-1} - x_n\rangle \geq -\left(\frac{1}{2}\|x_n - x_{n+1}\|^2 + \frac{1}{2}\|x_{n-1} - x_n\|^2\right),$$

and from (3. 69) we get

$$\frac{\sigma}{2}\|x_{n+1} - x_n\|^2 \leq D_F(x_{n+1}, x_n).$$

Hence, (3.77) leads to

$$(f+g)(x_{n+1}) + \frac{\sigma - L_{\nabla g}\alpha_n - \beta_n}{2\alpha_n}\|x_{n+1} - x_n\|^2 \\ \leq (f+g)(x_n) + \frac{\beta_n}{2\alpha_n}\|x_{n-1} - x_n\|^2. \qquad (3.\ 78)$$

Obviously $M_1 = \frac{\sigma - L_{\nabla g}\overline{\alpha}}{2\overline{\alpha}} - \frac{\beta}{2\underline{\alpha}} \leq \frac{\sigma - L_{\nabla g}\alpha_n - \beta_n}{2\alpha_n}$ and $M_2 = \frac{\beta}{2\underline{\alpha}} \geq \frac{\beta_n}{2\alpha_n}$ thus,

$$(f+g)(x_{n+1}) + M_1\|x_n - x_{n+1}\|^2 \leq (f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2$$

and the first part of the lemma is proved.

Let $0 < \underline{\alpha} \leq \overline{\alpha}$ and $\beta > 0$ be such that $\sigma > \overline{\alpha}L_{\nabla g} + 2\beta\frac{\overline{\alpha}}{\underline{\alpha}}$. One can immediately see that the latter is equivalent to $M_1 > M_2$ and the proof is complete. □

Remark 3.10 If $\underline{\alpha}$ and β are positive numbers such that $\sigma > \underline{\alpha}L_{\nabla g} + 2\beta$, then

$$\underline{\alpha} < \frac{\underline{\alpha}\sigma}{\underline{\alpha}L_{\nabla g} + 2\beta}.$$

By choosing

$$\underline{\alpha} \leq \overline{\alpha} < \frac{\underline{\alpha}\sigma}{\underline{\alpha}L_{\nabla g} + 2\beta},$$

relation (3. 76) is satisfied.

On the other hand, if $\overline{\alpha}$ and β are positive numbers such that $\sigma > \overline{\alpha}L_{\nabla g} + 2\beta$, then

$$\frac{2\beta\overline{\alpha}}{\sigma - \overline{\alpha}L_{\nabla g}} < \overline{\alpha}.$$

By choosing

$$\frac{2\beta\overline{\alpha}}{\sigma - \overline{\alpha}L_{\nabla g}} < \underline{\alpha} \leq \overline{\alpha},$$

relation (3. 76) is again satisfied.

Proposition 3.1 *In the setting of Problem 3.5, chose $\underline{\alpha}, \overline{\alpha}, \beta$ satisfying* (3. 76) *and M_1, M_2 satisfying* (3. 75). *Assume that $f+g$ is bounded from below. Then the following statements hold:*

(a) $\sum_{n\geq 1}\|x_n - x_{n-1}\|^2 < +\infty$;

(b) the sequence $((f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2)_{n\geq 1}$ is monotonically decreasing and convergent;

(c) the sequence $((f+g)(x_n))_{n\in\mathbb{N}}$ is convergent.

Proof. For every $n \geq 1$, set $a_n = (f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2$ and $b_n = (M_1 - M_2)\|x_n - x_{n+1}\|^2$. Then obviously from Lemma 3.6 one has for every $n \geq 1$

$$a_{n+1} + b_n = (f+g)(x_{n+1}) + M_1\|x_n - x_{n+1}\|^2 \leq (f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2 = a_n.$$

The conclusion follows now from Lemma 1.2. □

Lemma 3.7 *In the setting of Problem 3.5, consider the sequences generated by Algorithm 3.3. For every $n \geq 1$ we have*

$$y_{n+1} \in \partial(f+g)(x_{n+1}), \qquad (3.\ 79)$$

where

$$y_{n+1} = \frac{\nabla F(x_n) - \nabla F(x_{n+1})}{\alpha_n} + \nabla g(x_{n+1}) - \nabla g(x_n) + \frac{\beta_n}{\alpha_n}(x_n - x_{n-1}).$$

Moreover,

$$\|y_{n+1}\| \le \frac{L_{\nabla F} + \alpha_n L_{\nabla g}}{\alpha_n} \|x_n - x_{n+1}\| + \frac{\beta_n}{\alpha_n} \|x_n - x_{n-1}\| \ \forall n \ge 1. \tag{3. 80}$$

Proof. We fix $n \ge 1$. From (3. 71) we have that

$$\frac{\nabla F(x_n) - \nabla F(x_{n+1})}{\alpha_n} - \nabla g(x_n) + \frac{\beta_n}{\alpha_n}(x_n - x_{n-1}) \in \partial f(x_{n+1}),$$

or, equivalently,

$$y_{n+1} - \nabla g(x_{n+1}) \in \partial f(x_{n+1}),$$

which shows that $y_{n+1} \in \partial(f+g)(x_{n+1})$.

The inequality (3. 80) follows now from the definition of y_{n+1} and the triangle inequality. $\square$

Lemma 3.8 *In the setting of Problem 3.5, chose* $\underline{\alpha}, \overline{\alpha}, \beta$ *satisfying* (3. 76) *and* M_1, M_2 *satisfying* (3. 75). *Assume that* $f + g$ *is coercive, i.e.*

$$\lim_{\|x\| \to +\infty} (f+g)(x) = +\infty.$$

Then any sequence $(x_n)_{n \in \mathbb{N}}$ *generated by Algorithm 3.3 has a subsequence convergent to a critical point of* $f+g$. *Actually every cluster point of* $(x_n)_{n \in \mathbb{N}}$ *is a critical point of* $f+g$.

Proof. Since $f+g$ is a proper, lower semicontinuous and coercive function, it follows that $\inf_{x \in \mathbb{R}^m} [f(x) + g(x)]$ is finite and the infimum is attained. Hence $f+g$ is bounded from below.

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 3.3. According to Proposition 3.1(b), we have

$$\begin{aligned}(f+g)(x_n) &\le (f+g)(x_n) + M_2\|x_n - x_{n-1}\|^2 \\ &\le (f+g)(x_1) + M_2\|x_1 - x_0\|^2 \ \forall n \ge 1.\end{aligned}$$

Since the function $f+g$ is coercive, its lower level sets are bounded, thus the sequence $(x_n)_{n \in \mathbb{N}}$ is bounded.

Let x be a cluster point of $(x_n)_{n \in \mathbb{N}}$. Then there exists a subsequence $(x_{n_k})_{k \in \mathbb{N}}$ such that $x_{n_k} \to x$ as $k \to +\infty$. We show that $(f+g)(x_{n_k}) \to (f+g)(x)$ as $k \to +\infty$ and that x is a critical point of $f+g$, that is $0 \in \partial(f+g)(x)$.

We show first that $f(x_{n_k}) \to f(x)$ as $k \to +\infty$. Since f is lower semicontinuous one has

$$\liminf_{k \to +\infty} f(x_{n_k}) \ge f(x).$$

On the other hand, from (3. 70) we have for every $n \ge 1$

$$\begin{aligned}&D_F(x_{n+1}, x_n) + \alpha_n \langle x_{n+1}, \nabla g(x_n)\rangle + \beta_n \langle x_{n+1}, x_{n-1} - x_n\rangle + \alpha_n f(x_{n+1}) \\ &\qquad \le D_F(x, x_n) + \alpha_n \langle x, \nabla g(x_n)\rangle + \beta_n \langle x, x_{n-1} - x_n\rangle + \alpha_n f(x),\end{aligned}$$

which leads to

$$\begin{aligned}&\frac{1}{\alpha_{n_k-1}} \left(D_F(x_{n_k}, x_{n_k-1}) - D_F(x, x_{n_k-1})\right) + \\ &\frac{1}{\alpha_{n_k-1}} \left(\langle x_{n_k} - x, \alpha_{n_k-1}\nabla g(x_{n_k-1}) - \beta_{n_k-1}(x_{n_k-1} - x_{n_k-2})\rangle\right) + \\ &\qquad f(x_{n_k}) \le f(x) \ \forall k \ge 2.\end{aligned}$$

The latter combined with Proposition 3.1(a) and (3. 69) shows that

$$\limsup_{k\to+\infty} f(x_{n_k}) \le f(x),$$

hence $\lim_{k\to+\infty} f(x_{n_k}) = f(x)$. Since g is continuous, obviously $g(x_{n_k}) \to g(x)$ as $k \to +\infty$, thus $(f+g)(x_{n_k}) \to (f+g)(x)$ as $k \to +\infty$.

Further, by using the notations from Lemma 3.7, we have $y_{n_k} \in \partial(f+g)(x_{n_k})$ for every $k \ge 2$. By Proposition 3.1(a) and Lemma 3.7 we get $y_{n_k} \to 0$ as $k \to +\infty$.

Concluding, we have:

$$y_{n_k} \in \partial(f+g)(x_{n_k})\ \forall k \ge 2,$$

$$(x_{n_k}, y_{n_k}) \to (x, 0), \text{ as } k \to +\infty,$$

$$(f+g)(x_{n_k}) \to (f+g)(x), \text{ as } k \to +\infty.$$

Hence $0 \in \partial(f+g)(x)$, that is, x is a critical point of $f+g$. □

Lemma 3.9 *In the setting of Problem 3.5, chose $\underline{\alpha}, \overline{\alpha}, \beta$ satisfying* (3. 76) *and M_1, M_2 satisfying* (3. 75)*. Assume that $f+g$ is coercive and consider the function*

$$H : \mathbb{R}^m \times \mathbb{R}^m \to (-\infty, +\infty],\ H(x,y) = (f+g)(x) + M_2\|x-y\|^2\ \forall (x,y) \in \mathbb{R}^m \times \mathbb{R}^m.$$

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.3. Then there exist $M, N > 0$ such that the following statements hold:

(H_1) $H(x_{n+1}, x_n) + M\|x_{n+1} - x_n\|^2 \le H(x_n, x_{n-1})$ for all $n \ge 1$;

(H_2) for all $n \ge 1$, there exists $w_{n+1} \in \partial H(x_{n+1}, x_n)$ such that

$$\|w_{n+1}\| \le N(\|x_{n+1} - x_n\| + \|x_n - x_{n-1}\|);$$

(H_3) if $(x_{n_k})_{k\in\mathbb{N}}$ is a subsequence such that $x_{n_k} \to x$ as $k \to +\infty$, then we have $H(x_{n_k}, x_{n_k-1}) \to H(x,x)$ as $k \to +\infty$ (there exists at least one subsequence with this property).

Proof For (H_1) just take $M = M_1 - M_2$ and the conclusion follows from Lemma 3.6.

Let us prove (H_2). For every $n \ge 1$ we define

$$w_{n+1} = (y_{n+1} + 2M_2(x_{n+1} - x_n), 2M_2(x_n - x_{n+1})),$$

where $(y_n)_{n\ge2}$ is the sequence introduced in Lemma 3.7. The fact that $w_{n+1} \in \partial H(x_{n+1}, x_n)$ follows from Lemma 3.7 and the relation

$$\partial H(x,y) = \big(\partial(f+h)(x) + 2M_2(x-y)\big) \times \{2M_2(y-x)\}\ \forall (x,y) \in \mathbb{R}^m \times \mathbb{R}^m. \quad (3.\ 81)$$

Further, one has (see also Lemma 3.7)

$$\begin{aligned}\|w_{n+1}\| &\le \|y_{n+1} + 2M_2(x_{n+1} - x_n)\| + \|2M_2(x_n - x_{n+1})\| \\ &\le \left(\frac{L_{\nabla F}}{\alpha_n} + L_{\nabla g} + 4M_2\right)\|x_{n+1} - x_n\| + \frac{\beta_n}{\alpha_n}\|x_n - x_{n-1}\|.\end{aligned}$$

Since $0 < \underline{\alpha} \le \alpha_n \le \overline{\alpha}$ and $0 \le \beta_n \le \beta$ for all $n \ge 1$, one can chose

$$N = \sup_{n\ge1}\left\{\frac{L_{\nabla F}}{\alpha_n} + L_{\nabla g} + 4M_2, \frac{\beta_n}{\alpha_n}\right\} < +\infty$$

and the conclusion follows.

For (H_3), consider $(x_{n_k})_{k\in\mathbb{N}}$ a subsequence such that $x_{n_k} \to x$ as $k \to +\infty$. We have shown in the proof of Lemma 3.8 that $(f+g)(x_{n_k}) \to (f+g)(x)$ as $k \to +\infty$. From Proposition 3.1(a) and the definition of H we easily derive that $H(x_{n_k}, x_{n_k-1}) \to H(x,x) = (f+g)(x)$ as $k \to +\infty$. The existence of such a sequence follows from Lemma 3.8. □

In the following we denote by $\omega((x_n)_{n\in\mathbb{N}})$ the set of cluster points of the sequence $(x_n)_{n\in\mathbb{N}}$.

Lemma 3.10 *In the setting of Problem 3.5, chose* $\underline{\alpha}, \overline{\alpha}, \beta$ *satisfying* (3. 76) *and* M_1, M_2 *satisfying* (3. 75). *Assume that $f+g$ is coercive and consider the function*

$$H : \mathbb{R}^m \times \mathbb{R}^m \to (-\infty, +\infty],\ H(x,y) = (f+g)(x) + M_2\|x-y\|^2\ \forall (x,y) \in \mathbb{R}^m \times \mathbb{R}^m.$$

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.3. Then the following statements are true:

(a) $\omega((x_n, x_{n-1})_{n\geq 1}) \subseteq \operatorname{crit}(H) = \{(x,x) \in \mathbb{R}^m \times \mathbb{R}^m : x \in \operatorname{crit}(f+g)\}$*;*

(b) $\lim_{n\to\infty} \operatorname{dist}((x_n, x_{n-1}), \omega((x_n, x_{n-1}))_{n\geq 1}) = 0$*;*

(c) $\omega((x_n, x_{n-1})_{n\geq 1})$ *is nonempty, compact and connected;*

(d) H is finite and constant on $\omega((x_n, x_{n-1})_{n\geq 1})$.

Proof. (a) According to Lemma 3.8 and Proposition 3.1(a) we have

$$\omega((x_n, x_{n-1})_{n\geq 1}) \subseteq \{(x,x) \in \mathbb{R}^m \times \mathbb{R}^m : x \in \operatorname{crit}(f+g)\}.$$

The equality

$$\operatorname{crit}(H) = \{(x,x) \in \mathbb{R}^m \times \mathbb{R}^m : x \in \operatorname{crit}(f+g)\}$$

follows from (3. 81).

(b) and (c) can be shown as in [35, Lemma 5], by also taking into consideration [35, Remark 5], where it is noticed that the properties (b) and (c) are generic for sequences satisfying $x_{n+1} - x_n \to 0$ as $n \to +\infty$.

(d) According to Proposition 3.1, the sequence $((f+g)(x_n))_{n\in\mathbb{N}}$ is convergent, i.e.

$$\lim_{n\to+\infty} (f+g)(x_n) = l \in \mathbb{R}.$$

Take an arbitrary $(x,x) \in \omega((x_n, x_{n-1})_{n\geq 1})$, where $x \in \operatorname{crit}(f+g)$ (we took statement (a) into consideration). From Lemma 3.9(H_3) it follows that there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ such that $x_{n_k} \to x$ as $k \to +\infty$ and $H(x_{n_k}, x_{n_k-1}) \to H(x,x)$ as $k \to +\infty$. Moreover, from Proposition 3.1 one has

$$H(x,x) = \lim_{k\to+\infty} H(x_{n_k}, x_{n_k-1}) = \lim_{k\to+\infty} (f+g)(x_{n_k}) + M_2\|x_{n_k} - x_{n_k-1}\|^2 = l$$

and the conclusion follows. □

We give now the main result concerning the convergence of the whole sequence $(x_n)_{n\in\mathbb{N}}$.

Theorem 3.8 *In the setting of Problem 3.5, chose* $\underline{\alpha}, \overline{\alpha}, \beta$ *satisfying* (3. 76) *and* M_1, M_2 *satisfying* (3. 75). *Assume that $f+g$ is coercive and that*

$$H : \mathbb{R}^m \times \mathbb{R}^m \to (-\infty, +\infty],\ H(x,y) = (f+g)(x) + M_2\|x-y\|^2\ \forall (x,y) \in \mathbb{R}^m \times \mathbb{R}^m$$

is a KL function. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.3. Then the following statements are true:

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

(b) there exists $x \in \operatorname{crit}(f+g)$ *such that* $\lim_{n\to+\infty} x_n = x$.

Proof. (a) Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.3. According to Lemma 3.10 we can consider an element $\overline{x} \in \operatorname{crit}(f+g)$ such that $(\overline{x},\overline{x}) \in \omega((x_n, x_{n-1})_{n\geq 1})$. In analogy to the proof of Lemma 3.9 (by taking into account also the decrease property (H1)) one can easily show that

$$\lim_{n\to+\infty} H(x_n, x_{n-1}) = H(\overline{x},\overline{x}).$$

We separately treat the following two cases.

I. There exists $\overline{n} \in \mathbb{N}$ such that $H(x_{\overline{n}}, x_{\overline{n}-1}) = H(\overline{x},\overline{x})$. The decrease property $(H1)$ in Lemma 3.9 implies $H(x_n, x_{n-1}) = H(\overline{x},\overline{x})$ for every $n \geq \overline{n}$. By using again property $(H1)$ in Lemma 3.9, one can show inductively that the sequence $(x_n, x_{n-1})_{n\geq\overline{n}}$ is constant. From here the conclusion follows automatically.

II. For all $n \geq 1$ we have $H(x_n, x_{n-1}) > H(\overline{x},\overline{x})$. Take $\Omega := \omega((x_n, x_{n-1})_{n\geq 1})$.

In virtue of Lemma 3.10(c) and (d) and Lemma 3.4, the KL property of H leads to the existence of positive numbers ϵ and η and a concave function $\varphi \in \Theta_\eta$ such that for all

$$\begin{aligned}(x,y) \in &\{(u,v) \in \mathbb{R}^m \times \mathbb{R}^m : \operatorname{dist}((u,v),\Omega) < \epsilon\} \\ &\cap \{(u,v) \in \mathbb{R}^m \times \mathbb{R}^m : H(\overline{x},\overline{x}) < H(u,v) < H(\overline{x},\overline{x}) + \eta\}\end{aligned} \tag{3. 82}$$

one has

$$\varphi'(H(x,y) - H(\overline{x},\overline{x}))\operatorname{dist}((0,0), \partial H(x,y)) \geq 1. \tag{3. 83}$$

Let $n_1 \in \mathbb{N}$ such that $H(x_n, x_{n-1}) < H(\overline{x},\overline{x}) + \eta$ for all $n \geq n_1$. According to Lemma 3.10(b), there exists $n_2 \in \mathbb{N}$ such that $\operatorname{dist}((x_n, x_{n-1}), \Omega) < \epsilon$ for all $n \geq n_2$.

Hence the sequence $(x_n, x_{n-1})_{n\geq\overline{n}}$ where $\overline{n} = \max\{n_1, n_2\}$, belongs to the intersection (3. 82). So we have (see (3. 83))

$$\varphi'(H(x_n, x_{n-1}) - H(\overline{x},\overline{x}))\operatorname{dist}((0,0), \partial H(x_n, x_{n-1})) \geq 1 \ \forall n \geq \overline{n}.$$

Since φ is concave, it holds

$$\begin{aligned}\varphi(H(x_n, x_{n-1}) - H(\overline{x},\overline{x})) - \varphi(H(x_{n+1}, x_n) - H(\overline{x},\overline{x})) \geq \\ \varphi'(H(x_n, x_{n-1}) - H(\overline{x},\overline{x})) \cdot (H(x_n, x_{n-1}) - H(x_{n+1}, x_n)) \geq \\ \frac{H(x_n, x_{n-1}) - H(x_{n+1}, x_n)}{\operatorname{dist}((0,0), \partial H(x_n, x_{n-1}))} \ \forall n \geq \overline{n}.\end{aligned}$$

Let $M, N > 0$ be the real numbers furnished by Lemma 3.9. According to Lemma 3.9(H_2) there exists $w_n \in \partial H(x_n, x_{n-1})$ such that

$$\|w_n\| \leq N(\|x_n - x_{n-1}\| + \|x_{n-1} - x_{n-2}\|) \ \forall n \geq 2.$$

Then obviously $\operatorname{dist}((0,0), \partial H(x_n, x_{n-1})) \leq \|w_n\|$, hence

$$\begin{aligned}\varphi(H(x_n, x_{n-1}) - H(\overline{x},\overline{x})) - \varphi(H(x_{n+1}, x_n) - H(\overline{x},\overline{x})) \geq \\ \frac{H(x_n, x_{n-1}) - H(x_{n+1}, x_n)}{\|w_n\|} \geq \\ \frac{H(x_n, x_{n-1}) - H(x_{n+1}, x_n)}{N(\|x_n - x_{n-1}\| + \|x_{n-1} - x_{n-2}\|)} \ \forall n \geq \overline{n}.\end{aligned}$$

On the other hand, from Lemma 3.9(H_1) we obtain that

$$H(x_n, x_{n-1}) - H(x_{n+1}, x_n) \geq M\|x_{n+1} - x_n\|^2 \ \forall n \geq 1.$$

Hence, one has

$$\varphi(H(x_n,x_{n-1})-H(\overline{x},\overline{x}))-\varphi(H(x_{n+1},x_n)-H(\overline{x},\overline{x}))\geq \frac{M\|x_{n+1}-x_n\|^2}{N(\|x_n-x_{n-1}\|+\|x_{n-1}-x_{n-2}\|)}\ \forall n\geq\overline{n}.$$

For all $n\geq 1$, let us denote

$$\frac{N}{M}(\varphi(H(x_n,x_{n-1})-H(\overline{x},\overline{x}))-\varphi(H(x_{n+1},x_n)-H(\overline{x},\overline{x})))=\epsilon_n$$

and

$$\|x_n-x_{n-1}\|=a_n.$$

Then the last inequality becomes

$$\epsilon_n\geq\frac{a_{n+1}^2}{a_n+a_{n-1}}\ \forall n\geq\overline{n}. \tag{3. 84}$$

Obviously, since $\varphi\geq 0$, for $S\geq 1$ we have

$$\begin{aligned}\sum_{n=1}^{S}\epsilon_n&=\frac{N}{M}(\varphi(H(x_1,x_0)-H(\overline{x},\overline{x}))-\varphi(H(x_{S+1},x_S)-H(\overline{x},\overline{x})))\\&\leq\frac{N}{M}(\varphi(H(x_1,x_0)-H(\overline{x},\overline{x}))),\end{aligned}$$

hence $\sum_{n\geq 1}\epsilon_n<+\infty$.

On the other hand, from (3. 84) we derive

$$a_{n+1}=\sqrt{\epsilon_n(a_n+a_{n-1})}\leq\frac{1}{4}(a_n+a_{n-1})+\epsilon_n\ \forall n\geq\overline{n}.$$

Hence, according to Lemma 3.5, $\sum_{n\geq 1}a_n<+\infty$, that is $\sum_{n\in\mathbb{N}}\|x_n-x_{n+1}\|<+\infty$.

(b) It follows from (a) that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, hence it is convergent. Applying Lemma 3.8, there exists $x\in\operatorname{crit}(f+g)$ such that $\lim_{n\to+\infty}x_n=x$. □

Remark 3.11 A similar conclusion to the one of Theorem 3.8 can be obtained by applying [11, Theorem 2.9] in $\mathbb{R}^m\times\mathbb{R}^m$ endowed with the Euclidean product topology for the function

$$\widetilde{H}:\mathbb{R}^m\times\mathbb{R}^m\to\overline{\mathbb{R}},\widetilde{H}(x,y)=(f+g)(x)+\frac{1}{2}(M_1+M_2)\|x-y\|^2.$$

Indeed, a direct consequence of Lemma 3.6 is the following inequality which holds for all $n\geq 1$

$$\widetilde{H}(x_{n+1},x_n)+\frac{1}{2}(M_1-M_2)(\|x_{n+1}-x_n\|^2+\|x_n-x_{n-1}\|^2)\leq\widetilde{H}(x_n,x_{n-1}).$$

This shows that H1 in [11] is fulfilled. The assumptions H2 and H3 in the above-mentioned article are direct consequences of (H_2) and, respectively, (H_3) in Lemma 3.9. Under these considerations, provided that $\widetilde{H}$ is a KL function, one obtains via [11, Theorem 2.9] the same conclusion as in Theorem 3.8.

However, the hypothesis that H is a KL function, as assumed in Theorem 3.8, is in our opinion in this context the most natural one, at least in what concerns the way in which it approaches the non-inertial case. Indeed, if β is equal to zero, then M_2 is equal to zero, too, and the conclusion of Theorem 3.8 follows by only assuming that $f+g$ is a KL function. On the other hand, in order to apply [11, Theorem 2.9], one would ask that $(x,y)\mapsto(f+g)(x)+\frac{1}{2}M_1\|x-y\|^2$ is a KL function, which is in general a stronger assumption.

Since the class of semi-algebraic functions is closed under addition (see for example [35]) and $(x, y) \mapsto c\|x - y\|^2$ is semi-algebraic for $c > 0$, we obtain also the following direct consequence.

Corollary 3.1 *In the setting of Problem 3.5, chose* $\underline{\alpha}, \overline{\alpha}, \beta$ *satisfying* (3. 76) *and* M_1, M_2 *satisfying* (3. 75). *Assume that* $f + g$ *is coercive and semi-algebraic. Let* $(x_n)_{n\in\mathbb{N}}$ *be a sequence generated by Algorithm 3.3. Then the following statements are true:*

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

(b) there exists $x \in \operatorname{crit}(f + g)$ *such that* $\lim_{n\to+\infty} x_n = x$.

Remark 3.12 As one can notice by taking a closer look at the proof of Lemma 3.8, the conclusion of this statement as the ones of Lemma 3.9, Lemma 3.10, Theorem 3.8 and Corollary 3.1 remain true, if instead of imposing that $f + g$ is coercive, we assume that $f + g$ is bounded from below and the sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 3.3 is bounded. This observation is useful when dealing with optimization problems as the ones considered in Subsection 3.3.2.

3.3.2 Numerical experiments

This subsection is dedicated to the presentation of two numerical experiments which illustrate the applicability of the algorithm proposed in this work. In both numerical experiments we considered $F = \frac{1}{2}\|\cdot\|^2$ and set $\sigma = 1$.

Detecting minimizers of nonconvex optimization problems

As emphasized in [110, Section 5.1] and [30, Exercise 1.3.9] one of the aspects which makes algorithms with inertial/memory effects useful is given by the fact that they are able to detect optimal solutions of minimization problems which cannot be found by their non-inertial variants. In this subsection we show that this phenomenon arises even when solving problems of type (3. 85), where the nonsmooth function f is nonconvex. A similar situation has been addressed in [110], however, by assuming that f is convex. Consider the optimization problem

$$\inf_{(x_1,x_2)\in\mathbb{R}^2} |x_1| - |x_2| + x_1^2 - \log(1 + x_1^2) + x_2^2. \tag{3. 85}$$

The function

$$f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = |x_1| - |x_2|,$$

is nonconvex and continuous, the function

$$g : \mathbb{R}^2 \to \mathbb{R}, g(x_1, x_2) = x_1^2 - \log(1 + x_1^2) + x_2^2,$$

is continuously differentiable with Lipschitz continuous gradient with Lipschitz constant $L_{\nabla g} = 9/4$ and one can easily prove that $f + g$ is coercive. Furthermore, combining [10, the remarks after Definition 4.1], [33, Remark 5(iii)] and [35, Section 5: Example 4 and Theorem 3], one can easily conclude that H in Theorem 3.8 is a KL function. By considering the first order optimality conditions

$$-\nabla g(x_1, x_2) \in \partial f(x_1, x_2) = \partial(|\cdot|)(x_1) \times \partial(-|\cdot|)(x_2)$$

and by noticing that for all $x \in \mathbb{R}$ we have

$$\partial(|\cdot|)(x) = \begin{cases} 1, & \text{if } x > 0 \\ -1, & \text{if } x < 0 \\ \text{[-1,1]}, & \text{if } x = 0 \end{cases}$$

and

$$\partial(-|\cdot|)(x) = \begin{cases} -1, & \text{if } x > 0, \\ 1, & \text{if } x < 0, \\ \{-1, 1\}, & \text{if } x = 0, \end{cases}$$

(for the latter, see for example [101]), one can easily determine the two critical points $(0, 1/2)$ and $(0, -1/2)$ of (3. 85), which are actually both optimal solutions of this minimization problem. In Figure 3.4 the level sets and the graph of the objective function in (3. 85) are represented.

For $\gamma > 0$ and $x = (x_1, x_2) \in \mathbb{R}^2$ we have (see Remark 3.9)

$$\operatorname{prox}_{\gamma f}(x) = \underset{u \in \mathbb{R}^2}{\operatorname{argmin}} \left\{ \frac{\|u - x\|^2}{2\gamma} + f(u) \right\} = \operatorname{prox}_{\gamma|\cdot|}(x_1) \times \operatorname{prox}_{\gamma(-|\cdot|)}(x_2),$$

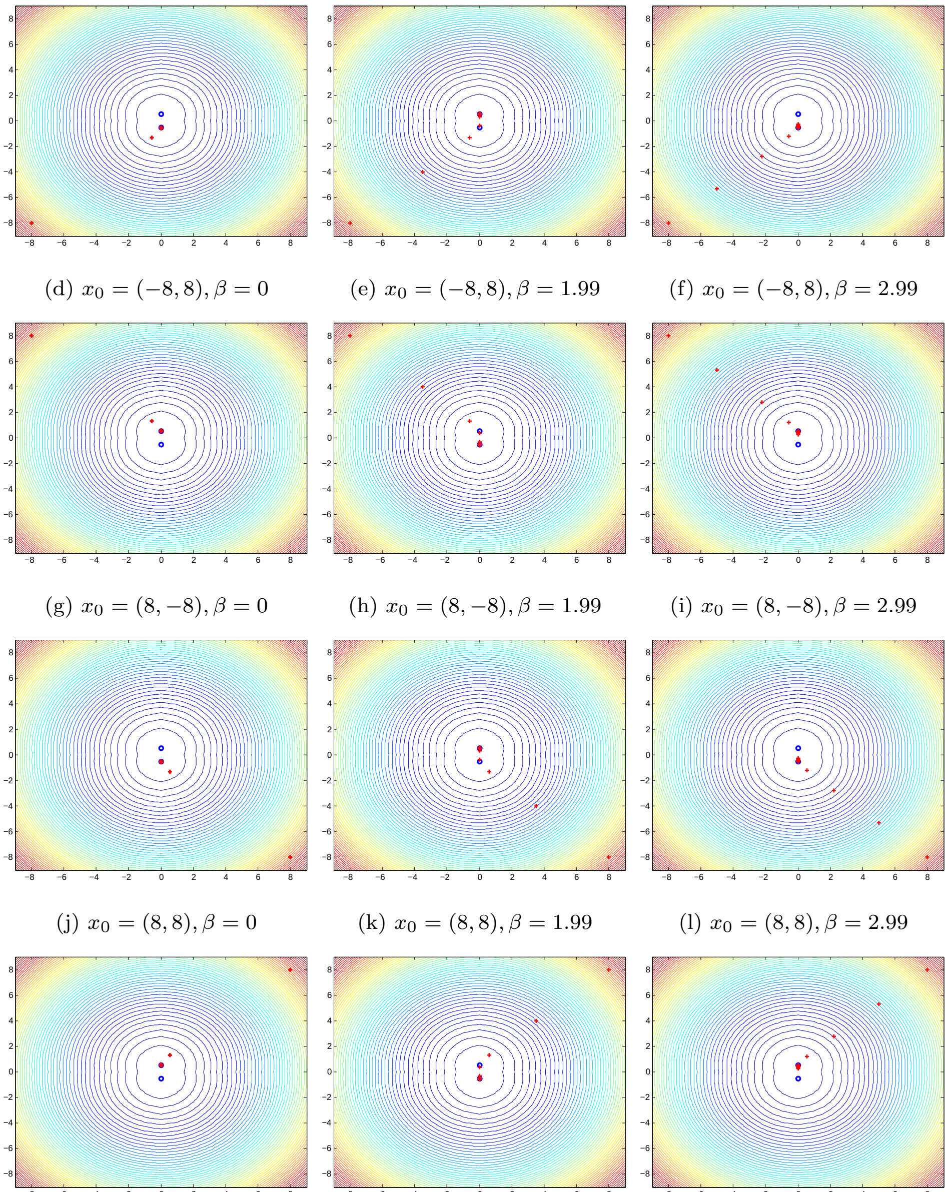
where in the first component one has the well-known shrinkage operator

$$\operatorname{prox}_{\gamma|\cdot|}(x_1) = x_1 - \operatorname{sgn}(x_1) \cdot \min\{|x_1|, \gamma\},$$

while for the proximal operator in the second component the following formula can be proven

$$\operatorname{prox}_{\gamma(-|\cdot|)}(x_2) = \begin{cases} x_2 + \gamma, & \text{if } x_2 > 0 \\ x_2 - \gamma, & \text{if } x_2 < 0 \\ \{-\gamma, \gamma\}, & \text{if } x_2 = 0. \end{cases}$$

We implemented Algorithm 3.3 by choosing $\beta_n = \beta = 0$ for all $n \geq 1$ (which corresponds to the non-inertial version), $\beta_n = \beta = 0.199$ for all $n \geq 1$ and $\beta_n = \beta = 0.299$ for all $n \geq 1$, respectively, and by setting $\alpha_n = (0.99999 - 2\beta_n)/L_{\nabla g}$ for all $n \geq 1$. As starting points we considered the corners of the box generated by the points $(\pm 8, \pm 8)$. Figure 3.3 shows that independently of the four starting points we have the following phenomenon: the non-inertial version recovers only one of the two optimal solutions, situation which persists even when changing the value of α_n; on the other hand, the inertial version is capable to find both optimal solutions, namely, one for $\beta = 0.199$ and the other one for $\beta = 0.299$.



(a) $x_0 = (-8, -8), \beta = 0$ (b) $x_0 = (-8, -8), \beta = 1.99$ (c) $x_0 = (-8, -8), \beta = 2.99$

(d) $x_0 = (-8, 8), \beta = 0$ (e) $x_0 = (-8, 8), \beta = 1.99$ (f) $x_0 = (-8, 8), \beta = 2.99$

(g) $x_0 = (8, -8), \beta = 0$ (h) $x_0 = (8, -8), \beta = 1.99$ (i) $x_0 = (8, -8), \beta = 2.99$

(j) $x_0 = (8, 8), \beta = 0$ (k) $x_0 = (8, 8), \beta = 1.99$ (l) $x_0 = (8, 8), \beta = 2.99$

Figure 3.3: Algorithm 3.3 after 100 iterations and with starting points $(-8, -8), (-8, 8), (8, -8)$ and $(8, 8)$, respectively: the first column shows the iterates of the non-inertial version ($\beta_n = \beta = 0$ for all $n \geq 1$), the second column the ones of the inertial version with $\beta_n = \beta = 1.99$ for all $n \geq 1$ and the third column the ones of the inertial version with $\beta_n = \beta = 2.99$ for all $n \geq 1$.

Restoration of noisy blurred images

The following numerical experiment concerns the restoration of a noisy blurred image by using a nonconvex misfit functional with nonconvex regularization. For a given matrix $A \in \mathbb{R}^{m \times m}$ describing a blur operator and a given vector $b \in \mathbb{R}^m$ representing the blurred and noisy image, the task is to estimate the unknown original image $\overline{x} \in \mathbb{R}^m$ fulfilling

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

(a) Contour plot (b) Graph

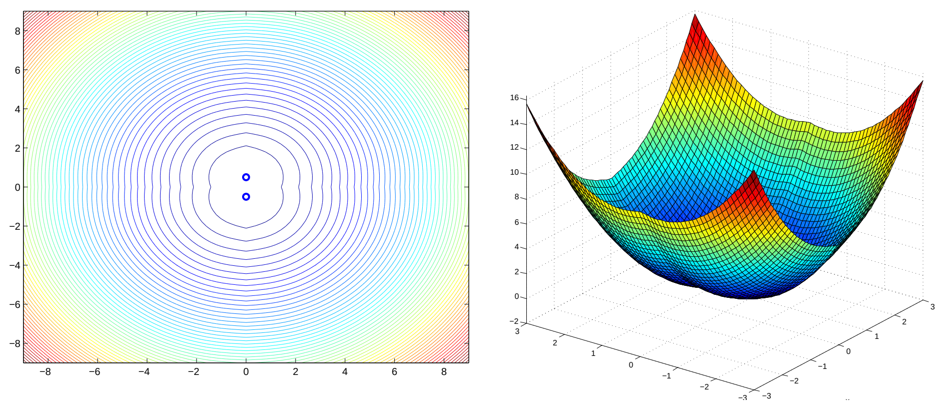


Figure 3.4: Contour plot and graph of the objective function in (3. 85). The two global optimal solutions $(0, 0.5)$ and $(0, -0.5)$ are marked on the first image.

To this end we solve the following regularized nonconvex minimization problem

$$\inf_{x \in \mathbb{R}^m} \left\{ \sum_{k=1}^{M} \sum_{l=1}^{N} \varphi\big((Ax-b)_{kl}\big) + \lambda \|Wx\|_0 \right\}, \tag{3. 86}$$

where $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\varphi(t) = \log(1+t^2),$$

is derived form the Student's t distribution, $\lambda > 0$ is a regularization parameter, $W : \mathbb{R}^m \to \mathbb{R}^m$ is a discrete Haar wavelet transform with four levels and

$$\|y\|_0 = \sum_{i=1}^{m} |y_i|_0$$

($|\cdot|_0 = |\operatorname{sgn}(\cdot)|$) furnishes the number of nonzero entries of a given vector $y = (y_1, ..., y_m) \in \mathbb{R}^m$. In this context, $x \in \mathbb{R}^m$ represents the vectorized image $X \in \mathbb{R}^{M \times N}$, where $m = M \cdot N$ and $x_{i,j}$ denotes the normalized value of the pixel located in the i-th row and the j-th column, for $i = 1, \ldots, M$ and $j = 1, \ldots, N$. Again, by combining [10, the remarks after Definition 4.1], [33, Remark 5(iii)] and [35, Section 5: Example 3, Example 4 and Theorem 3], one can conclude that H in Theorem 3.8 is a KL function.

It is immediate that (3. 86) can be written in the form (3. 68), by defining

$$f(x) = \lambda \|Wx\|_0$$

and

$$g(x) = \sum_{k=1}^{M} \sum_{l=1}^{N} \varphi\big((Ax-b)_{kl}\big)$$

for all $x \in \mathbb{R}^m$. By using that $WW^* = W^*W = I_m$, one can prove the following formula concerning the proximal operator of f

$$\operatorname{prox}_{\gamma f}(x) = W^* \operatorname{prox}_{\lambda\gamma\|\cdot\|_0}(Wx) \ \forall x \in \mathbb{R}^m \ \forall \gamma > 0,$$

where for all $u = (u_1, ..., u_m)$ we have (see [11, Example 5.4(a)])

$$\operatorname{prox}_{\lambda\gamma\|\cdot\|_0}(u) = \big(\operatorname{prox}_{\lambda\gamma|\cdot|_0}(u_1), ..., \operatorname{prox}_{\lambda\gamma|\cdot|_0}(u_m)\big)$$

and for all $t \in \mathbb{R}$

$$\operatorname{prox}_{\lambda\gamma|\cdot|_0}(t) = \begin{cases} t, & \text{if } |t| > \sqrt{2\lambda\gamma}, \\ \{0, t\}, & \text{if } |t| = \sqrt{2\lambda\gamma}, \\ 0, & \text{otherwise.} \end{cases}$$

For the experiments we used the 256×256 boat test image which we first blurred by using a Gaussian blur operator of size 9×9 and standard deviation 4 and to which we afterward added a zero-mean white Gaussian noise with standard deviation 10^{-6}. In the first row of Figure 3.5 the original boat test image and the blurred and noisy one are represented, while in the second row one has the reconstructed images by means of the non-inertial (for $\beta_n = \beta = 0$ for all $n \geq 1$) and inertial versions (for $\beta_n = \beta = 10^{-7}$ for all $n \geq 1$) of Algorithm 1, respectively. We took as regularization parameter $\lambda = 10^{-5}$ and set $\alpha_n = (0.999999 - 2\beta_n)/L_{\nabla g}$ for all $n \geq 1$, whereby the Lipschitz constant of the gradient of the smooth misfit function is $L_{\nabla g} = 2$.

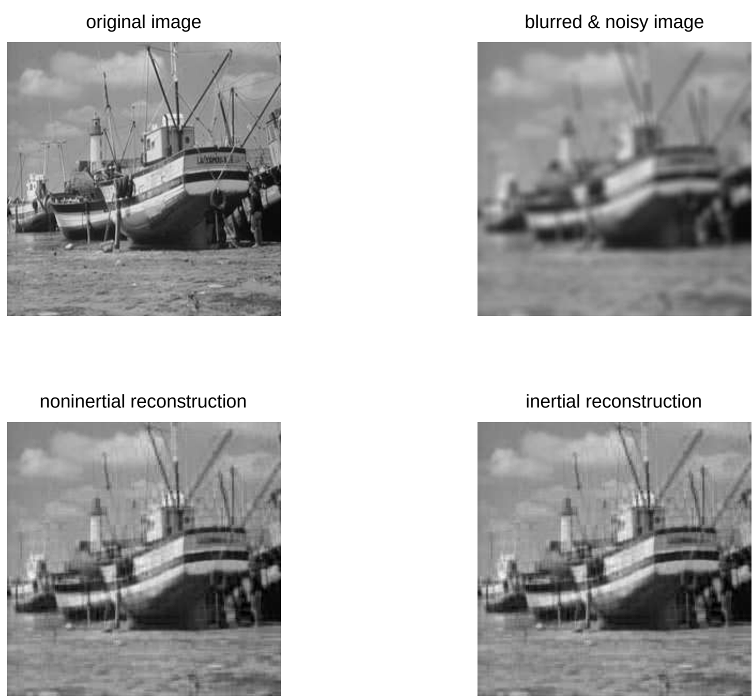


Figure 3.5: The first row shows the original 256×256 boat test image and the blurred and noisy one and the second row the reconstructed images after 300 iterations.

We compared the quality of the recovered images for $\beta_n = \beta$ for all $n \geq 1$ and different values of β by making use of the improvement in signal-to-noise ratio (ISNR), which is defined as

$$\mathrm{ISNR}(n) = 10 \log_{10} \left(\frac{\|x - b\|^2}{\|x - x_n\|^2} \right),$$

where x, b and x_n denote the original, observed and estimated image at iteration n, respectively.

β	0.4	0.2	0.01	0.0001	10^{-7}	0
ISNR(300)	2.081946	3.101028	3.492989	3.499428	3.511135	3.511134

Table 3.3: The ISNR values after 300 iterations for different choices of β.

In Table 3.3 we list the values of the ISNR-function after 300 iterations, whereby the case $\beta = 0$ corresponds to the non-inertial version of the algorithm. One can notice that for β taking very small values, the inertial version is competitive with the non-inertial one.

Chapter 4

Penalty-type splitting algorithms for monotone inclusion problems

It is the aim of this chapter to present and investigate penalty-type methods for monotone inclusion problems. In Section 4.1 we pay attention on forward-backward-type penalty methods, while in Section 4.2 we consider Tseng's type penalty schemes for monotone inclusion problems, including highly structured inclusions involving composition with linear and continuous operators and parallel-sums.

We need some additional notions and technical results which are recalled in the following.

The *Fitzpatrick function* associated to a monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$, defined as

$$\varphi_A : \mathcal{H} \times \mathcal{H} \to \overline{\mathbb{R}},\ \varphi_A(x,u) = \sup_{(y,v)\in \mathrm{gr}\, A} \{\langle x,v\rangle + \langle y,u\rangle - \langle y,v\rangle\},$$

is a convex and lower semicontinuous function and it will play an important role throughout this chapter. Let us note that a similar object has been considered also by Krylov in 1982, see [93]. The terminology used in the literature is *Fitzpatrick function*, due to [87], where some fundamental properties have been investigated in connection with monotone operators. Let us underline that this notion opened the gate towards the employment of convex analysis specific tools when investigating the maximality of monotone operators (see [26, 27, 36–39, 64, 124] and the references therein). In case A is maximally monotone, φ_A is proper and it fulfills

$$\varphi_A(x,u) \geq \langle x,u\rangle\ \forall (x,u) \in \mathcal{H} \times \mathcal{H},$$

with equality if and only if $(x,u) \in \mathrm{gr}\, A$. Notice that if $f : \mathcal{H} \to \overline{\mathbb{R}}$, is a proper, convex and lower semi-continuous function, then the following inequality is true (see [27])

$$\varphi_{\partial f}(x,u) \leq f(x) + f^*(u)\ \forall (x,u) \in \mathcal{H} \times \mathcal{H}. \tag{4. 1}$$

We refer the reader to [27], for formulae of the corresponding Fitzpatrick functions computed for particular classes of monotone operators.

The following ergodic version of the Opial Lemma will be used several times in this chapter. Let $(x_n)_{n\geq 1}$ be a sequence in $\mathcal{H}$ and $(\lambda_k)_{k\geq 1}$ a sequence of positive numbers such that $\sum_{k\geq 1} \lambda_k = +\infty$. Let $(z_n)_{n\geq 1}$ be the sequence of weighted averages defined as (see [16])

$$z_n = \frac{1}{\tau_n} \sum_{k=1}^{n} \lambda_k x_k,\ \text{where } \tau_n = \sum_{k=1}^{n} \lambda_k\ \forall n \geq 1. \tag{4. 2}$$

Lemma 4.1 *(Opial-Passty, see [112,113] and [15, Lemma 2.1]) Let C be a nonempty subset of $\mathcal{H}$ and assume that the limes $\lim_{n\to\infty}\|x_n - x\|$ exists for every $x \in C$. If every weak sequential cluster point of $(x_n)_{n\geq 1}$ (respectively $(z_n)_{n\geq 1}$) lies in C, then $(x_n)_{n\geq 1}$ (respectively $(z_n)_{n\geq 1}$) converges weakly to an element in C as $n \to +\infty$.*

4.1 A forward-backward penalty scheme

In this section we propose and investigate the convergence properties of a forward-backward penalty type scheme for solving inclusion problems governed by monotone operators. The problem we deal with at the beginning of this section has the following formulation.

Problem 4.1 *Let $\mathcal{H}$ be a real Hilbert space, $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ maximally monotone operators, $D : \mathcal{H} \to \mathcal{H}$ an η-cocoercive operator with $\eta > 0$ and suppose that $M = \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Dx + N_M(x).$$

The following iterative scheme for solving Problem 4.1 is inspired by [16].

Algorithm 4.1

```
Initialization:  Choose x_1 ∈ H
     For n ≥ 1:  Choose w_n ∈ Bx_n
                 Set x_{n+1} = J_{λ_n A}(x_n − λ_n Dx_n − λ_n β_n w_n),
```

where $(\lambda_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ are sequences of positive real numbers. Notice that Algorithm 4.1 is well-defined, if $\operatorname{dom} B = \mathcal{H}$, which will be the case in the next section, when B is assumed to be cocoercive. For the convergence statement the following hypotheses are needed

$$(H_{fitz}) \begin{cases} (i)\ A + N_M \text{ is maximally monotone and } \operatorname{zer}(A + D + N_M) \neq \emptyset; \\ (ii)\ \text{For every } p \in \operatorname{ran} N_M : \\ \qquad \sum_{n\geq 1} \lambda_n\beta_n \left[\sup\limits_{u\in M} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] < +\infty; \\ (iii)\ (\lambda_n)_{n\geq 1} \in \ell^2 \setminus \ell^1. \end{cases}$$

Remark 4.1 Since A is maximally monotone and M is a nonempty convex and closed set, $A + N_M$ is maximally monotone if a so-called regularity condition is fulfilled. This is the cae if one of the Rockafellar conditions $M \cap \operatorname{int}\operatorname{dom} A \neq \emptyset$ or $\operatorname{dom} A \cap \operatorname{int} M \neq \emptyset$, is fulfilled (see [121]). We refer the reader to [26,36–39,124,131] for further conditions which guarantee the maximality of the sum of maximally monotone operators. Further, we refer to [26, Subsection 23.4] for conditions eunsuring that the set of zeros of a maximally monotone operator is nonempty.

Further, as D is maximally monotone (see [26, Example 20.28]) and $\operatorname{dom} D = \mathcal{H}$, the hypothesis (i) above guarantees that $A + D + N_M$ is maximally monotone, too (see [26, Corollary 24.4]). Moreover, for each $p \in \operatorname{ran} N_M$ we have

$$\sup_{u\in M} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right) \geq 0 \ \forall n \geq 1.$$

Indeed, if $p \in \operatorname{ran} N_M$, then there exists $\overline{u} \in M$ such that $p \in N_M(\overline{u})$. This implies that

$$\sup_{u\in M} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right) \geq \left\langle \overline{u}, \frac{p}{\beta_n}\right\rangle - \sigma_M\left(\frac{p}{\beta_n}\right) = 0 \ \forall n \geq 1.$$

Remark 4.2 Let us underline that the hypothesis (ii) is a generalization of the condition considered in [16] (we refer to (H_{fitz}^{opt}) and Remark 4.7 in Section 4.2.3 for conditions guaranteeing (ii)). Indeed, if $Dx = 0$ for all $x \in \mathcal{H}$ and $B = \partial\Psi$, where $\Psi : \mathcal{H} \to \overline{\mathbb{R}}$ is a proper, convex and lower semicontinuous function with $\min \Psi = 0$, then the monotone inclusion in Problem 4.1 becomes

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + N_M(x), \tag{4. 3}$$

since in this case $M = \operatorname{argmin} \Psi$. This problem has been investigated in [16] under the condition

$$(H) \begin{cases} (i)\ A + N_M \text{ is maximally monotone and } \operatorname{zer}(A + N_M) \neq \emptyset; \\ \quad \text{For every } p \in \operatorname{ran} N_M : \\ \qquad \sum_{n\geq 1} \lambda_n\beta_n \left[\Psi^*\left(\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] < +\infty; \\ (iii)\ (\lambda_n)_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1. \end{cases}$$

Moreover, as $\Psi(x) = 0$ for all $x \in M$, by (4. 1) it follows that condition (ii) in (H) implies condition (ii) in (H_{fitz}), hence the hypothesis formulated by means of the Fitzpatrick function extends the one given [16] to the more general setting considered in Problem 4.1. It remains an open question to find examples of proper, convex and lower semicontinuous functions $\Psi : \mathcal{H} \to \overline{\mathbb{R}}$ with $\min \Psi = 0$ for which (ii) in (H) is not fulfilled, while for $B = \partial\Psi$ condition (ii) in (H_{fitz}) holds.

4.1.1 The general case

In this subsection we will prove an abstract convergence result for Algorithm 4.1, which will be subsequently refined in the case when B is a cocoercive operator. Some techniques from [16] are adapted to the more general setting we consider here.

Lemma 4.2 *Let $(x_n)_{n\geq 1}$ and $(w_n)_{n\geq 1}$ be the sequences generated by Algorithm 4.1 and take $(u, w) \in \operatorname{gr}(A + D + N_M)$ such that $w = v + p + Du$, where $v \in Au$ and $p \in N_M(u)$. Then the following inequality holds for all $n \geq 1$*

$$\begin{aligned} &\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \lambda_n(2\eta - 3\lambda_n)\|Dx_n - Du\|^2 \\ &\qquad \leq 2\lambda_n\beta_n \left[\sup_{u\in M} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] \\ &\qquad\quad + 3\lambda_n^2\beta_n^2\|w_n\|^2 + 3\lambda_n^2\|Du + v\|^2 + 2\lambda_n\langle u - x_n, w\rangle. \end{aligned} \tag{4. 4}$$

Proof. From the definition of the resolvent of A we have

$$\frac{x_n - x_{n+1}}{\lambda_n} - \beta_n w_n - Dx_n \in Ax_{n+1}$$

and since $v \in Au$, the monotonicity of A guarantees

$$\langle x_{n+1} - u, x_n - x_{n+1} - \lambda_n(\beta_n w_n + Dx_n + v)\rangle \geq 0\ \forall n \geq 1, \tag{4. 5}$$

thus

$$\langle u - x_{n+1}, x_n - x_{n+1}\rangle \leq \lambda_n\langle u - x_{n+1}, \beta_n w_n + Dx_n + v\rangle\ \forall n \geq 1.$$

Further, since

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

we get for any $n \geq 1$

$$\begin{aligned}
&\|x_{n+1}-u\|^2 - \|x_n-u\|^2 \\
\leq\ & 2\lambda_n\langle u-x_{n+1}, \beta_n w_n + Dx_n + v\rangle - \|x_{n+1}-x_n\|^2 \\
=\ & 2\lambda_n\langle u-x_n, \beta_n w_n + Dx_n + v\rangle + 2\lambda_n\langle x_n - x_{n+1}, \beta_n w_n + Dx_n + v\rangle \\
& - \|x_{n+1}-x_n\|^2 \\
\leq\ & 2\lambda_n\langle u-x_n, \beta_n w_n + Dx_n + v\rangle + \lambda_n^2\|\beta_n w_n + Dx_n + v\|^2 \\
\leq\ & 2\lambda_n\langle u-x_n, \beta_n w_n + Dx_n + v\rangle + 3\lambda_n^2\beta_n^2\|w_n\|^2 + 3\lambda_n^2\|Du+v\|^2 \\
& + 3\lambda_n^2\|Dx_n - Du\|^2.
\end{aligned}$$

Next we evaluate the first term on the right hand-side of the last of the above inequalities. By using the cocoercivity of D and the definition of the Fitzpatrick function and that $w_n \in Bx_n$ and $\sigma_M\left(\frac{p}{\beta_n}\right) = \langle u, \frac{p}{\beta_n}\rangle$ for every $n \geq 1$, we obtain

$$\begin{aligned}
& 2\lambda_n\langle u-x_n, \beta_n w_n + Dx_n + v\rangle \\
=\ & 2\lambda_n\langle u-x_n, \beta_n w_n + Dx_n + w - p - Du\rangle \\
=\ & 2\lambda_n\langle u-x_n, Dx_n - Du\rangle + 2\lambda_n\langle u-x_n, \beta_n w_n - p\rangle + 2\lambda_n\langle u-x_n, w\rangle \\
=\ & 2\lambda_n\langle u-x_n, Dx_n - Du\rangle + 2\lambda_n\beta_n\left(\langle u, w_n\rangle + \left\langle x_n, \frac{p}{\beta_n}\right\rangle - \langle x_n, w_n\rangle - \left\langle u, \frac{p}{\beta_n}\right\rangle\right) \\
& + 2\lambda_n\langle u-x_n, w\rangle \\
\leq\ & -2\eta\lambda_n\|Dx_n - Du\|^2 \\
& + 2\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n\langle u-x_n, w\rangle.
\end{aligned}$$

This provides the desired conclusion. □

Theorem 4.1 *Let $(x_n)_{n\geq 1}$ and $(w_n)_{n\geq 1}$ be the sequences generated by Algorithm 4.1 and $(z_n)_{n\geq 1}$ the sequence defined in (4. 2). If (H_{fitz}) is fulfilled and the condition $(\lambda_n\beta_n\|w_n\|)_{n\geq 1} \in \ell^2$ holds, then $(z_n)_{n\geq 1}$ converges weakly to an element in* $\operatorname{zer}(A+D+N_M)$ *as $n \to +\infty$.*

Proof. As $\lim_{n\to+\infty}\lambda_n = 0$, there exists $n_0 \in \mathbb{N}$ such that $2\eta - 3\lambda_n \geq 0$ for all $n \geq n_0$. Thus, for $(u, w) \in \operatorname{gr}(A+D+N_M)$, such that $w = v + p + Du$, where $v \in Au$ and $p \in N_M(u)$, by (4. 4) it holds for all $n \geq n_0$

$$\begin{aligned}
\|x_{n+1}-u\|^2 - \|x_n-u\|^2 \leq\ & 2\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] \\
& + 3\lambda_n^2\beta_n^2\|w_n\|^2 + 3\lambda_n^2\|Du+v\|^2 + 2\lambda_n\langle u-x_n, w\rangle.
\end{aligned} \tag{4. 6}$$

By Lemma 4.1, it is sufficient to prove that the following two statements hold:

(a) for every $u \in \operatorname{zer}(A+D+N_M)$ the sequence $(\|x_n - u\|)_{n\geq 1}$ is convergent;

(b) every weak sequential cluster point of $(z_n)_{n\geq 1}$ lies in $\operatorname{zer}(A+D+N_M)$.

(a) For every $u \in \operatorname{zer}(A+D+N_M)$ one can take $w = 0$ in (4. 6) and the conclusion follows from Lemma 1.2.

(b) Let z be a weak sequential cluster point of $(z_n)_{n\geq 1}$. As we already noticed that $A+D+N_M$ is maximally monotone, in order to show that $z \in \operatorname{zer}(A+D+N_M)$ we will use the characterization given in (1. 29). Take $(u, w) \in \operatorname{gr}(A+D+N_M)$

such that $w = v + p + Du$, where $v \in Au$ and $p \in N_M(u)$. Let be $N \in \mathbb{N}$ with $N \geq n_0 + 2$. Summing up for $n = n_0 + 1, ..., N$ the inequalities in (4. 6), we get

$$\|x_{N+1}-u\|^2-\|x_{n_0+1}-u\|^2 \leq L+2\left\langle \sum_{n=1}^{N}\lambda_n u - \sum_{n=1}^{N}\lambda_n x_n - \sum_{n=1}^{n_0}\lambda_n u + \sum_{n=1}^{n_0}\lambda_n x_n, w \right\rangle,$$

where

$$\begin{aligned} L = 2 &\sum_{n\geq n_0+1}\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_M\left(\frac{p}{\beta_n}\right)\right] \\ &+ \ 3\sum_{n\geq n_0+1}\lambda_n^2\beta_n^2\|w_n\|^2 + 3\sum_{n\geq n_0+1}\lambda_n^2\|Du+v\|^2 \in \mathbb{R}. \end{aligned}$$

Discarding the nonnegative term $\|x_{N+1}-u\|^2$ and dividing by $2\tau_N = 2\sum_{k=1}^{N}\lambda_k$ we obtain

$$-\frac{\|x_{n_0+1}-u\|^2}{2\tau_N} \leq \frac{\widetilde{L}}{2\tau_N} + \langle u - z_N, w\rangle,$$

where $\widetilde{L} := L + 2\langle -\sum_{n=1}^{n_0}\lambda_n u + \sum_{n=1}^{n_0}\lambda_n x_n, w\rangle \in \mathbb{R}$. By passing to the limit as $N \to +\infty$ and using that $\lim_{N\to+\infty}\tau_N = +\infty$, we get

$$\liminf_{N\to+\infty}\langle u - z_N, w\rangle \geq 0.$$

Since z is a weak sequential cluster point of $(z_n)_{n\geq 1}$, we obtain that $\langle u - z, w\rangle \geq 0$. Finally, as this inequality holds for arbitrary $(u, w) \in \mathrm{gr}(A + D + N_M)$, the desired conclusion follows. $\square$

In the following we show that strong monotonicity of the operator A ensures strong convergence of the sequence $(x_n)_{n\geq 1}$.

Theorem 4.2 *Let $(x_n)_{n\geq 1}$ and $(w_n)_{n\geq 1}$ be the sequences generated by Algorithm 4.1. If (H_{fitz}) is fulfilled, $(\lambda_n\beta_n\|w_n\|)_{n\geq 1} \in \ell^2$ and the operator A is γ-strongly monotone with $\gamma > 0$, then $(x_n)_{n\geq 1}$ converges strongly to the unique element in* $\mathrm{zer}(A + D + N_M)$ *as $n \to +\infty$.*

Proof. Let be $u \in \mathrm{zer}(A + D + N_M)$ and $w = 0 = v + p + Du$, where $v \in Au$ and $p \in N_M(u)$. Since A is γ-strongly monotone, inequality (4. 5) becomes

$$\langle x_{n+1} - u, x_n - x_{n+1} - \lambda_n(\beta_n w_n + Dx_n + v)\rangle \geq \lambda_n\gamma\|x_{n+1} - u\|^2 \ \forall n \geq 1. \quad (4.\ 7)$$

Following the lines of the proof of Lemma 4.2 for $w = 0$ we obtain for all $n \geq 1$

$$\begin{aligned} &2\gamma\lambda_n\|x_{n+1}-u\|^2 + \|x_{n+1}-u\|^2 - \|x_n-u\|^2 + \lambda_n(2\eta-3\lambda_n)\|Dx_n-Du\|^2 \\ &\leq 2\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_M\left(\frac{p}{\beta_n}\right)\right] + 3\lambda_n^2\beta_n^2\|w_n\|^2 + 3\lambda_n^2\|Du+v\|^2. \end{aligned}$$

Thus, as $\lim_{n\to+\infty}\lambda_n = 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$\begin{aligned} &2\gamma\lambda_n\|x_{n+1}-u\|^2 + \|x_{n+1}-u\|^2 - \|x_n-u\|^2 \\ \leq\ &2\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_M\left(\frac{p}{\beta_n}\right)\right] + 3\lambda_n^2\beta_n^2\|w_n\|^2 + 3\lambda_n^2\|Du+v\|^2 \end{aligned}$$

and, so,

$$
\begin{aligned}
2\gamma \sum_{n \geq n_0} \lambda_n \|x_{n+1} - u\|^2 \leq\ & \|x_{n_0} - u\|^2 \\
& + 2 \sum_{n \geq n_0} \lambda_n \beta_n \left[\sup_{u \in M} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_M \left(\frac{p}{\beta_n} \right) \right] \\
& + 3 \sum_{n \geq n_0} \lambda_n^2 \beta_n^2 \|w_n\|^2 + 3\|Du + v\|^2 \sum_{n \geq n_0} \lambda_n^2 \\
<\ & +\infty.
\end{aligned}
$$

Since $\sum_{n \geq 1} \lambda_n = +\infty$ and $(\|x_n - u\|)_{n \geq 1}$ is convergent (see the proof of Theorem 4.1 (a)), it follows $\lim_{n \to +\infty} \|x_n - u\| = 0$. $\square$

4.1.2 The case B is cocoercive

In this subsection we deal with the situation when B is a (single-valued) cocoercive operator. Our aim is to show that in this setting the assumption $(\lambda_n \beta_n \|w_n\|)_{n \geq 1} \in \ell^2$ in Theorem 4.1 and Theorem 4.2 can be replaced by a milder condition involving only the sequences $(\lambda_n)_{n \geq 1}$ and $(\beta_n)_{n \geq 1}$. The problem under consideration has the following formulation.

Problem 4.2 *Let $\mathcal{H}$ be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ an η-cocoercive operator with $\eta > 0$, $B : \mathcal{H} \to \mathcal{H}$ a μ-cocoercive operator with $\mu > 0$ and suppose that $M = \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is*

$$\textit{find } x \in \mathcal{H} \textit{ such that } 0 \in Ax + Dx + N_M(x).$$

Algorithm 4.1 has in this particular setting the following formulation.

Algorithm 4.2
`Initialization:` `Choose` $x_1 \in \mathcal{H}$
`For` $n \geq 1$ `set:` $x_{n+1} = J_{\lambda_n A}(x_n - \lambda_n D x_n - \lambda_n \beta_n B x_n)$.

Remark 4.3 (a) If $Dx = 0$ for every $x \in \mathcal{H}$ and $B = \nabla \Psi$, where $\Psi : \mathcal{H} \to \mathbb{R}$ is a convex and differentiable function with μ^{-1}-Lipschitz continuous gradient for $\mu > 0$ fulfilling $\min \Psi = 0$, then we rediscover the setting considered in [16, Section 3], while Algorithm 4.2 becomes the iterative method investigated in that paper.

(b) In case $Bx = 0$ for all $x \in \mathcal{H}$ Algorithm 4.2 turns out to be classical forward-backward scheme (see [26,75,130]), since under these premises $M = \mathcal{H}$, hence $N_M(x) = \{0\}$ for all $x \in \mathcal{H}$.

Before stating the convergence result for Algorithm 4.2 some technical results are in order.

Lemma 4.3 *Let be $u \in M \cap \operatorname{dom} A$ and $v \in Au$. Then for every $\varepsilon \geq 0$ and all $n \geq 1$ we have*

$$
\begin{aligned}
& \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \frac{\varepsilon}{1+\varepsilon} \|x_{n+1} - x_n\|^2 + \frac{2\varepsilon}{1+\varepsilon} \lambda_n \beta_n \langle x_n - u, Bx_n \rangle \\
& \leq \lambda_n \beta_n \left((1+\varepsilon) \lambda_n \beta_n - \frac{2\mu}{1+\varepsilon} \right) \|Bx_n\|^2 + 2\lambda_n \langle u - x_{n+1}, Dx_n + v \rangle. \qquad (4.\ 8)
\end{aligned}
$$

Proof. As in the proof of Lemma 4.2 we obtain for all $n \geq 1$ that

$$\begin{aligned}
&\|x_{n+1}-u\|^2 - \|x_n-u\|^2 + \|x_{n+1}-x_n\|^2 \\
&\leq 2\lambda_n\langle u-x_{n+1}, \beta_n Bx_n + Dx_n + v\rangle \\
&= 2\lambda_n\beta_n\langle u-x_n, Bx_n\rangle + 2\lambda_n\beta_n\langle x_n-x_{n+1}, Bx_n\rangle \\
&\quad + 2\lambda_n\langle u-x_{n+1}, Dx_n+v\rangle.
\end{aligned}$$

Since B is μ-cocoercive and $Bu = 0$ we have that

$$\langle u-x_n, Bx_n\rangle \leq -\mu\|Bx_n\|^2 \ \forall n \geq 1,$$

hence for all $n \geq 1$ and $\varepsilon \geq 0$ it holds

$$2\lambda_n\beta_n\langle u-x_n, Bx_n\rangle \leq -\frac{2\mu}{1+\varepsilon}\lambda_n\beta_n\|Bx_n\|^2 + \frac{2\varepsilon}{1+\varepsilon}\lambda_n\beta_n\langle u-x_n, Bx_n\rangle.$$

Inequality (4. 8) follows by taking into consideration also that for all $n \geq 1$ and $\varepsilon \geq 0$ we have

$$2\lambda_n\beta_n\langle x_n-x_{n+1}, Bx_n\rangle \leq \frac{1}{1+\varepsilon}\|x_{n+1}-x_n\|^2 + (1+\varepsilon)\lambda_n^2\beta_n^2\|Bx_n\|^2.$$

□

Lemma 4.4 *Assume that* $\limsup_{n\to+\infty}\lambda_n\beta_n < 2\mu$ *and let be* $u \in M \cap \operatorname{dom} A$ *and* $v \in Au$*. Then there exist* $a, b > 0$ *and* $n_0 \in \mathbb{N}$ *such that for all* $n \geq n_0$ *it holds*

$$\begin{aligned}
&\|x_{n+1}-u\|^2 - \|x_n-u\|^2 + a\left(\|x_{n+1}-x_n\|^2 + \lambda_n\beta_n\langle x_n-u, Bx_n\rangle + \lambda_n\beta_n\|Bx_n\|^2\right) \\
&\leq \left(b\lambda_n^2 - 2\eta\lambda_n\right)\|Dx_n-Du\|^2 + 2\lambda_n\langle u-x_n, v+Du\rangle + b\lambda_n^2\|Du+v\|^2. \qquad (4.\ 9)
\end{aligned}$$

Proof. We start by noticing that, by making use of the cocoercivity of D, for every $\varepsilon > 0$ and all $n \geq 1$ it holds

$$\begin{aligned}
&2\lambda_n\langle u-x_{n+1}, Dx_n+v\rangle \\
&= 2\lambda_n\langle x_n-x_{n+1}, Dx_n+v\rangle + 2\lambda_n\langle u-x_n, Dx_n+v\rangle \\
&\leq \frac{\varepsilon}{2(1+\varepsilon)}\|x_{n+1}-x_n\|^2 + \frac{2(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Dx_n+v\|^2 + 2\lambda_n\langle u-x_n, Dx_n+v\rangle \\
&\leq \frac{\varepsilon}{2(1+\varepsilon)}\|x_{n+1}-x_n\|^2 + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Dx_n-Du\|^2 + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Du+v\|^2 \\
&\quad + 2\lambda_n\langle u-x_n, Dx_n-Du\rangle + 2\lambda_n\langle u-x_n, v+Du\rangle \\
&\leq \frac{\varepsilon}{2(1+\varepsilon)}\|x_{n+1}-x_n\|^2 + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Dx_n-Du\|^2 + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Du+v\|^2 \\
&\quad - 2\lambda_n\eta\|Dx_n-Du\|^2 + 2\lambda_n\langle u-x_n, v+Du\rangle.
\end{aligned}$$

In combination with (4. 8) it yields for every $\varepsilon > 0$ and every $n \geq 1$

$$\begin{aligned}
&\|x_{n+1}-u\|^2 - \|x_n-u\|^2 + \frac{\varepsilon}{2(1+\varepsilon)}\|x_{n+1}-x_n\|^2 + \frac{2\varepsilon}{1+\varepsilon}\lambda_n\beta_n\langle x_n-u, Bx_n\rangle \\
&\quad + \frac{\varepsilon}{1+\varepsilon}\lambda_n\beta_n\|Bx_n\|^2 \\
&\leq \lambda_n\beta_n\left((1+\varepsilon)\lambda_n\beta_n - \frac{2\mu}{1+\varepsilon} + \frac{\varepsilon}{1+\varepsilon}\right)\|Bx_n\|^2 \\
&\quad + \left(\frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2 - 2\eta\lambda_n\right)\|Dx_n-Du\|^2 \\
&\quad + 2\lambda_n\langle u-x_n, v+Du\rangle + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Du+v\|^2.
\end{aligned}$$

Since $\limsup_{n\to+\infty} \lambda_n\beta_n < 2\mu$, there exists $\alpha > 0$ and $n_0 \in \mathbb{N}$ such that $\lambda_n\beta_n < \alpha < 2\mu$ for all $n \geq n_0$. Hence, for all $n \geq n_0$ and every $\varepsilon > 0$ it holds

$$\lambda_n\beta_n\left((1+\varepsilon)\lambda_n\beta_n - \frac{2\mu}{1+\varepsilon} + \frac{\varepsilon}{1+\varepsilon}\right) < \alpha\left((1+\varepsilon)\alpha - \frac{2\mu}{1+\varepsilon} + \frac{\varepsilon}{1+\varepsilon}\right)$$

and one can take $\varepsilon_0 > 0$ small enough such that $(1+\varepsilon_0)\alpha - \frac{2\mu}{1+\varepsilon_0} + \frac{\varepsilon_0}{1+\varepsilon_0} < 0$. The desired conclusion follows by choosing $a = \frac{\varepsilon_0}{2(1+\varepsilon_0)}$ and $b = \frac{4(1+\varepsilon_0)}{\varepsilon_0}$. □

Lemma 4.5 *Assume that* $\limsup_{n\to+\infty} \lambda_n\beta_n < 2\mu$ *and* $\lim_{n\to+\infty} \lambda_n = 0$ *and let be* $(u,w) \in \operatorname{gr}(A+D+N_M)$ *such that* $w = v+p+Du$, *where* $v \in Au$ *and* $p \in N_M(u)$. *Then there exist* $a, b > 0$ *and* $n_1 \in \mathbb{N}$ *such that for all* $n \geq n_1$ *it holds*

$$\begin{aligned}
&\|x_{n+1}-u\|^2 - \|x_n-u\|^2 \qquad (4.\ 10)\\
&+ a\left(\|x_{n+1}-x_n\|^2 + \frac{\lambda_n\beta_n}{2}\langle x_n-u, Bx_n\rangle + \lambda_n\beta_n\|Bx_n\|^2\right)\\
\leq\ & \frac{a\lambda_n\beta_n}{2}\left[\sup_{u\in M}\varphi_B\left(u,\frac{4p}{a\beta_n}\right) - \sigma_M\left(\frac{4p}{a\beta_n}\right)\right] + 2\lambda_n\langle u-x_n, w\rangle + b\lambda_n^2\|Du+v\|^2. \qquad (4.\ 11)
\end{aligned}$$

Proof. According to Lemma 4.4, there exist $a, b > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ inequality (4. 9) holds. Since $\lim_{n\to\infty} \lambda_n = 0$, there exists $n_1 \in \mathbb{N}, n_1 \geq n_0$ such that $b\lambda_n^2 - 2\eta\lambda_n \leq 0$ for all $n \geq n_1$, hence,

$$\begin{aligned}
&\|x_{n+1}-u\|^2 - \|x_n-u\|^2\\
&+ a\left(\|x_{n+1}-x_n\|^2 + \lambda_n\beta_n\langle x_n-u, Bx_n\rangle + \lambda_n\beta_n\|Bx_n\|^2\right)\\
\leq\ & 2\lambda_n\langle u-x_n, v+Du\rangle + b\lambda_n^2\|Du+v\|^2\ \forall n \geq n_1.
\end{aligned}$$

The conclusion follows by combining this inequality with the subsequent estimation that holds for all $n \geq 1$:

$$\begin{aligned}
&2\lambda_n\langle u-x_n, v+Du\rangle + \frac{a\lambda_n\beta_n}{2}\langle u-x_n, Bx_n\rangle\\
=\ & 2\lambda_n\langle u-x_n, -p\rangle + \frac{a\lambda_n\beta_n}{2}\langle u-x_n, Bx_n\rangle + 2\lambda_n\langle u-x_n, w\rangle\\
=\ & \frac{a\lambda_n\beta_n}{2}\left(\langle u, Bx_n\rangle + \left\langle x_n, \frac{4p}{a\beta_n}\right\rangle - \langle x_n, Bx_n\rangle - \left\langle u, \frac{4p}{a\beta_n}\right\rangle\right) + 2\lambda_n\langle u-x_n, w\rangle\\
\leq\ & \frac{a\lambda_n\beta_n}{2}\left[\sup_{u\in M}\varphi_B\left(u,\frac{4p}{a\beta_n}\right) - \sigma_M\left(\frac{4p}{a\beta_n}\right)\right] + 2\lambda_n\langle u-x_n, w\rangle.
\end{aligned}$$

□

Theorem 4.3 *Let* $(x_n)_{n\geq1}$ *and* $(w_n)_{n\geq1}$ *be the sequences generated by Algorithm 4.2 and* $(z_n)_{n\geq1}$ *be the sequence defined in* (4. 2). *If* (H_{fitz}) *is fulfilled and the condition*

$$\limsup_{n\to+\infty} \lambda_n\beta_n < 2\mu$$

holds, then the following statements are true:

(i) for every $u \in \operatorname{zer}(A+D+N_M)$ *the sequence* $(\|x_n-u\|)_{n\geq1}$ *is convergent and the series* $\sum_{n\geq1}\|x_{n+1}-x_n\|^2$, $\sum_{n\geq1}\lambda_n\beta_n\langle Bx_n, x_n-u\rangle$ *and* $\sum_{n\geq1}\lambda_n\beta_n\|Bx_n\|^2$ *are convergent as well. In particular* $\lim_{n\to+\infty}\|x_{n+1}-x_n\| = 0$. *If, moreover,* $\liminf_{n\to+\infty}\lambda_n\beta_n > 0$, *then* $\lim_{n\to+\infty}\langle Bx_n, x_n-u\rangle = \lim_{n\to+\infty}\|Bx_n\| = 0$ *and every weak sequential cluster point of* $(x_n)_{n\geq1}$ *lies in* M.

(ii) $(z_n)_{n\geq 1}$ *converges weakly to an element in* $\operatorname{zer}(A + D + N_M)$ *as* $n \to +\infty$.

(iii) *if, additionally,* A *is strongly monotone, then* $(x_n)_{n\geq 1}$ *converges strongly to the unique element in* $\operatorname{zer}(A + D + N_M)$ *as* $n \to +\infty$.

Proof. For every $u \in \operatorname{zer}(A + D + N_M)$, according to Lemma 4.5, there exist $a, b > 0$ and $n_1 \in \mathbb{N}$ such that for all $n \geq n_1$ inequality (4. 10) is true for $w = 0$. This gives rise via Lemma 1.2 to the statements in (i). As the sequence $(\lambda_n\beta_n)_{n\geq 1}$ is bounded above, it automatically follows that $(\lambda_n\beta_n\|Bx_n\|)_{n\geq 1} \in \ell^2$. Hence, (ii) and (iii) follow as consequences of Theorem 4.1 and Theorem 4.2, respectively. □

Remark 4.4 We emphasize the fact that the results obtained in this subsection by assuming that B is a cocoercive operator enables us to treat the more general case where M is the set of zeros of an arbitrary maximally monotone operator.

Indeed, we consider in Problem 4.1 that $M = \operatorname{zer} N \neq \emptyset$, where $N : \mathcal{H} \rightrightarrows \mathcal{H}$ is a (possibly set-valued) maximally monotone operator. The idea is to apply the results in Subsection 2.1.2 to the operator $B := J_{N^{-1}} : \mathcal{H} \to \mathcal{H}$, which according to [26, Proposition 20.22, Corollary 23.10 and Proposition 4.2] is μ-cocoercive with $\mu = 1$. By noticing that $\operatorname{zer} J_{N^{-1}} = \operatorname{zer} N$, we can address the monotone inclusion problem to be solved as a problem formulated in the framework of Problem 4.2. Obviously, in the iterative scheme given in Algorithm 4.2 the operator N will be evaluated by a backward step.

Further, we will show that one can provide sufficient conditions for (ii) in (H_{fitz}) written in terms of the Fitzpatrick function of the operator N. To this end we use the following estimation of for Fitzpatrick function of $J_{N^{-1}}$, obtained by applying [27, Proposition 4.2], which is a result that gives an upper bound for the Fitzpatrick function of the sum of two maximally monotone operators in terms of the Fitzpatrick functions of the operators involved. Take an arbitrary $p \in \operatorname{ran} N_M$ and $u \in M$. We have for every $n \geq 1$

$$\begin{aligned}
\varphi_{J_{N^{-1}}}\left(u, \frac{p}{\beta_n}\right) &= \varphi_{\mathrm{Id}+N^{-1}}\left(\frac{p}{\beta_n}, u\right)\\
&\leq \varphi_{\mathrm{Id}}\left(\frac{p}{\beta_n}, 0\right) + \varphi_{N^{-1}}\left(\frac{p}{\beta_n}, u\right)\\
&= \frac{1}{4}\left\|\frac{p}{\beta_n}\right\|^2 + \varphi_N\left(u, \frac{p}{\beta_n}\right),
\end{aligned}$$

where we used the fact that $\varphi_{\mathrm{Id}}(x, v) = \frac{1}{4}\|x + v\|^2$ for all $(x, v) \in \mathcal{H} \times \mathcal{H}$.

This means that the condition (ii) in (H_{fitz}) applied to the reformulation of Problem 4.1 described above is fulfilled, if we assume that $\sum_{n\geq 1} \frac{\lambda_n}{\beta_n} < +\infty$ and that for every $p \in \operatorname{ran} N_M$, $\sum_{n\geq 1} \lambda_n\beta_n \left[\sup_{u\in M} \varphi_N\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] < +\infty$.

4.2 Tseng's Type penalty schemes

In this section we deal first with the monotone inclusion problem stated in Problem 4.2 by relaxing the cocoercivity of B and D to monotonicity and Lipschitz continuity. The iterative method we propose in this setting is a forward-backward-forward penalty scheme and relies on a method introduced by Tseng in [129] (see [26, 62, 76] for further details and motivations). By making use of primal-dual techniques we will be able then to employ the proposed approach when solving monotone inclusion problems involving parallel sums and compositions of maximally monotone operators with linear continuous ones.

4.2.1 Relaxing cocoercivity to monotonicity and Lipschitz continuity

We deal first with the following problem.

Problem 4.3 *Let $\mathcal{H}$ be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ a monotone and η^{-1}-Lipschitz continuous operator with $\eta > 0$, $B : \mathcal{H} \to \mathcal{H}$ a monotone and μ^{-1}-Lipschitz continuous operator with $\mu > 0$ and suppose that $M = \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Dx + N_M(x).$$

The investigated algorithm has the following form.

Algorithm 4.3

$$\begin{array}{rl} \texttt{Initialization}: & \texttt{Choose } x_1 \in \mathcal{H} \\ \texttt{For } n \geq 1 \texttt{ set}: & p_n = J_{\lambda_n A}(x_n - \lambda_n Dx_n - \lambda_n \beta_n Bx_n) \\ & x_{n+1} = \lambda_n \beta_n (Bx_n - Bp_n) + \lambda_n (Dx_n - Dp_n) + p_n, \end{array}$$

where $(\lambda_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ are sequences of positive real numbers.

Remark 4.5 If $Bx = 0$ for every $x \in \mathcal{H}$ (which corresponds to the situation $N_M(x) = \{0\}$ for all $x \in \mathcal{H}$), then Algorithm 4.3 turns out to be the error-free forward-backward-forward scheme from [62, Theorem 2.5] (see also [129]).

We start with the following technical statement.

Lemma 4.6 *Let $(x_n)_{n\geq 1}$ and $(p_n)_{n\geq 1}$ be the sequences generated by Algorithm 4.3 and let be $(u, w) \in \operatorname{gr}(A + D + N_M)$ such that $w = v + p + Du$, where $v \in Au$ and $p \in N_M(u)$. Then the following inequality holds for all $n \geq 1$:*

$$\begin{aligned} &\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(1 - \left(\frac{\lambda_n \beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\right) \|x_n - p_n\|^2 \\ &\leq 2\lambda_n \beta_n \left[\sup_{u \in M} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n \langle u - p_n, w\rangle. \end{aligned} \tag{4. 12}$$

Proof. From the definition of the resolvent we have

$$\frac{x_n - p_n}{\lambda_n} - \beta_n Bx_n - Dx_n \in Ap_n \ \forall n \geq 1$$

and since $v \in Au$, the monotonicity of A guarantees

$$\langle p_n - u, x_n - p_n - \lambda_n(\beta_n Bx_n + Dx_n + v)\rangle \geq 0 \ \forall n \geq 1,$$

thus

$$\langle u - p_n, x_n - p_n\rangle \leq \langle u - p_n, \lambda_n \beta_n Bx_n + \lambda_n Dx_n + \lambda_n v\rangle \ \forall n \geq 1.$$

By using the definition of x_{n+1} given in the algorithm we obtain

$$\begin{aligned} &\langle u - p_n, x_n - p_n\rangle \\ &\quad \leq \langle u - p_n, x_{n+1} - p_n + \lambda_n \beta_n Bp_n + \lambda_n Dp_n + \lambda_n v\rangle \\ &\quad = \langle u - p_n, x_{n+1} - p_n\rangle + \lambda_n \beta_n \langle u - p_n, Bp_n\rangle \\ &\qquad + \lambda_n \langle u - p_n, Dp_n\rangle + \lambda_n \langle u - p_n, v\rangle \ \forall n \geq 1. \end{aligned}$$

From here it follows for all $n \geq 1$

$$\begin{aligned}
&\frac{1}{2}\|u-p_n\|^2 - \frac{1}{2}\|x_n-u\|^2 + \frac{1}{2}\|x_n-p_n\|^2 \\
&\qquad \leq \frac{1}{2}\|u-p_n\|^2 - \frac{1}{2}\|x_{n+1}-u\|^2 + \frac{1}{2}\|x_{n+1}-p_n\|^2 \\
&\qquad\quad + \lambda_n\beta_n\langle u-p_n, Bp_n\rangle + \lambda_n\langle u-p_n, Dp_n\rangle + \lambda_n\langle u-p_n, v\rangle.
\end{aligned}$$

Since $v = w - p - Du$ and due to the fact that D is monotone, we obtain for every $n \geq 1$

$$\begin{aligned}
&\|x_{n+1}-u\|^2 - \|x_n-u\|^2 \\
&\qquad \leq \|x_{n+1}-p_n\|^2 - \|x_n-p_n\|^2 \\
&\qquad\quad + 2\lambda_n\beta_n\left(\langle u, Bp_n\rangle + \left\langle p_n, \frac{p}{\beta_n}\right\rangle - \langle p_n, Bp_n\rangle - \left\langle u, \frac{p}{\beta_n}\right\rangle\right) \\
&\qquad\quad + 2\lambda_n\langle u-p_n, Dp_n - Du\rangle + 2\lambda_n\langle u-p_n, w\rangle \\
&\qquad \leq \|x_{n+1}-p_n\|^2 - \|x_n-p_n\|^2 \\
&\qquad\quad + 2\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n\langle u-p_n, w\rangle.
\end{aligned}$$

The conclusion follows, by noticing that the Lipschitz continuity of B and D yields

$$\begin{aligned}
\|x_{n+1}-p_n\| &\leq \frac{\lambda_n\beta_n}{\mu}\|x_n-p_n\| + \frac{\lambda_n}{\eta}\|x_n-p_n\| \\
&= \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)\|x_n-p_n\| \ \forall n \geq 1.
\end{aligned}$$

□

The convergence of Algorithm 4.3 is stated below.

Theorem 4.4 *Let $(x_n)_{n\geq 1}$ and $(p_n)_{n\geq 1}$ be the sequences generated by Algorithm 4.3 and $(z_n)_{n\geq 1}$ the sequence defined in* (4. 2). *If (H_{fitz}) is fulfilled and the condition*

$$\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right) < 1$$

holds, then $(z_n)_{n\geq 1}$ converges weakly to an element in $\operatorname{zer}(A+D+N_M)$ *as $n \to +\infty$.*

Proof. The proof of the theorem relies on the following three statements:

(a) for every $u \in \operatorname{zer}(A+D+N_M)$ the sequence $(\|x_n-u\|)_{n\geq 1}$ is convergent;

(b) every weak cluster point of $(z'_n)_{n\geq 1}$, where

$$z'_n = \frac{1}{\tau_n}\sum_{k=1}^{n}\lambda_k p_k \text{ and } \tau_n = \sum_{k=1}^{n}\lambda_k \ \forall n \geq 1,$$

lies in $\operatorname{zer}(A+D+N_M)$;

(c) every weak cluster point of $(z_n)_{n\geq 1}$ lies in $\operatorname{zer}(A+D+N_M)$.

In order to show (a) and (b) one has only to slightly adapt the proof of Theorem 4.1 and this is why we omit to give further details. For (c) it is enough to prove that $\lim_{n\to+\infty}\|z_n - z'_n\| = 0$ and the statement of the theorem will be a consequence of Lemma 4.1.

Taking $u \in \operatorname{zer}(A + D + N_M)$ and $w = 0 = v + p + Du$, where $v \in Au$ and $p \in N_M(u)$, from (4. 12) we have

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(1 - \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\right)\|x_n - p_n\|^2$$
$$\leq 2\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right].$$

As $\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right) < 1$, we obtain by Lemma 1.2 that

$$\sum_{n\geq 1}\|x_n - p_n\|^2 < +\infty.$$

Moreover, for all $n \geq 1$ it holds

$$\|z_n - z_n'\|^2 = \frac{1}{\tau_n^2}\left\|\sum_{k=1}^{n}\lambda_k(x_k - p_k)\right\|^2 \leq \frac{1}{\tau_n^2}\left(\sum_{k=1}^{n}\lambda_k\|x_k - p_k\|\right)^2$$
$$\leq \frac{1}{\tau_n^2}\left(\sum_{k=1}^{n}\lambda_k^2\right)\left(\sum_{k=1}^{n}\|x_k - p_k\|^2\right).$$

Since $(\lambda_n)_{n\geq 1} \in \ell^2\setminus\ell^1$, by taking into consideration that $\tau_n = \sum_{k=1}^{n}\lambda_k \to +\infty$ $(n \to +\infty)$, we obtain $\|z_n - z_n'\| \to 0$ $(n \to +\infty)$. □

As it happens for the forward-backward penalty scheme, strong monotonicity of the operator A ensures strong convergence of the sequence $(x_n)_{n\geq 1}$.

Theorem 4.5 *Let $(x_n)_{n\geq 1}$ and $(p_n)_{n\geq 1}$ be the sequences generated by Algorithm 4.3. If (H_{fitz}) is fulfilled,*

$$\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right) < 1$$

and the operator A is γ-strongly monotone with $\gamma > 0$, then $(x_n)_{n\geq 1}$ converges strongly to the unique element in $\operatorname{zer}(A + D + N_M)$ as $n \to +\infty$.

Proof. Let be $u \in \operatorname{zer}(A + D + N_M)$ and $w = 0 = v + p + Du$, where $v \in Au$ and $p \in N_M(u)$. Following the lines of the proof of Lemma 4.6 one can easily show that

$$2\gamma\lambda_n\|p_n - u\|^2 + \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(1 - \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\right)\|x_n - p_n\|^2$$
$$\leq 2\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] \ \forall n \geq 1.$$

The hypotheses imply the existence of $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$

$$2\gamma\lambda_n\|p_n - u\|^2 + \|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq 2\lambda_n\beta_n\left[\sup_{u\in M}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right].$$

As in the proof of Theorem 4.2, from here it follows that

$$\sum_{n\geq 1}\lambda_n\|p_n - u\|^2 < +\infty.$$

Since $(\lambda_n)_{n\geq 1}$ is bounded above and $\sum_{n\in N} \|x_n - p_n\|^2 < +\infty$ (see the proof of Theorem 4.4), it yields

$$\sum_{n=1}^{\infty} \lambda_n \|x_n - u\|^2 \leq 2\sum_{n=1}^{\infty} \lambda_n \|x_n - p_n\|^2 + 2\sum_{n=1}^{\infty} \lambda_n \|p_n - u\|^2 < +\infty.$$

As $\sum_{n\geq 1} \lambda_n = +\infty$ and $(\|x_n - u\|)_{n\geq 1}$ is convergent, it follows $\lim_{n\to+\infty} \|x_n - u\| = 0$. $\square$

4.2.2 Primal-dual Tseng's type penalty schemes

In this section we propose a forward-backward-forward-type algorithm for solving the following monotone inclusion problem involving linearly composed and parallel-sum type monotone operators and investigate its convergence.

Problem 4.4 *Let $\mathcal{H}$ be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and $C : \mathcal{H} \to \mathcal{H}$ a monotone and ν-Lipschitz continuous operator for $\nu > 0$. Let m be a strictly positive integer and for every $i \in \{1, ..., m\}$ let $\mathcal{G}_i$ be a real Hilbert space, $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ a maximally monotone operator, $D_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ a monotone operator such that D_i^{-1} is ν_i-Lipschtz continuous for $\nu_i > 0$ and $L_i : \mathcal{H} \to \mathcal{G}_i$ a nonzero linear continuous operator. Consider also $B : \mathcal{H} \to \mathcal{H}$ a monotone and μ^{-1}-Lipschitz continuous operator with $\mu > 0$ and suppose that $M = \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is*

$$\textit{find } x \in \mathcal{H} \textit{ such that } 0 \in Ax + \sum_{i=1}^{m} L_i^*(B_i \Box D_i)(L_i x) + Cx + N_M(x). \qquad (4.\ 13)$$

Let us present our algorithm for solving this problem.

Algorithm 4.4

$$\begin{array}{rl}
\texttt{Initialization}: & \texttt{Choose } (x_1, v_{1,1}, ..., v_{m,1}) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m \\
\texttt{For } n \geq 1 \texttt{ set}: & p_n = J_{\lambda_n A}[x_n - \lambda_n(Cx_n + \sum_{i=1}^m L_i^* v_{i,n}) - \lambda_n\beta_n Bx_n] \\
& q_{i,n} = J_{\lambda_n B_i^{-1}}[v_{i,n} + \lambda_n(L_i x_n - D_i^{-1} v_{i,n})],\ i = 1, ..., m \\
& x_{n+1} = \lambda_n\beta_n(Bx_n - Bp_n) + \lambda_n(Cx_n - Cp_n) \\
& \qquad +\lambda_n \sum_{i=1}^m L_i^*(v_{i,n} - q_{i,n}) + p_n \\
& v_{i,n+1} = \lambda_n L_i(p_n - x_n) + \lambda_n(D_i^{-1} v_{i,n} - D_i^{-1} q_{i,n}) \\
& \qquad + q_{i,n},\ i = 1, ..., m,
\end{array}$$

where $(\lambda_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ are sequences of positive real numbers.

Remark 4.6 In case $Bx = 0$ for all $x \in \mathcal{H}$, Algorithm 4.4 collapses into the error-free variant of the iterative scheme given in [76, Theorem 3.1] for solving the monotone inclusion problem

$$0 \in Ax + \sum_{i=1}^{m} L_i^*(B_i \Box D_i)(L_i x) + Cx,$$

since in this case $M = \mathcal{H}$, hence $N_M(x) = \{0\}$ for all $x \in \mathcal{H}$.

For the convergence result we need the following additionally hypotheses (we refer the reader to the remarks 4.1 and 4.7 for sufficient conditions guaranteeing $(H_{fitz}^{par-sum})$):

$$(H_{fitz}^{par-sum}) \left\{ \begin{array}{l} (i)\ A+N_M \text{ is maximally monotone and} \\ \quad \operatorname{zer}\left(A+\sum_{i=1}^m L_i^* \circ (B_i \Box D_i) \circ L_i + C + N_M\right) \neq \emptyset; \\ (ii) \text{ For every } p \in \operatorname{ran} N_M: \\ \qquad \sum_{n\geq 1} \lambda_n \beta_n \left[\sup_{u\in M} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] < +\infty; \\ (iii)\ (\lambda_n)_{n\geq 1} \in \ell^2 \setminus \ell^1. \end{array} \right.$$

Let us give the main statement of this section. The proof relies on the fact that Problem 4.4 can be written in the same form as Problem 4.3 in an appropriate product space.

Theorem 4.6 *Consider the sequences generated by Algorithm 4.4 and $(z_n)_{n\geq 1}$ the sequence defined in (4. 2). Assume that $(H_{fitz}^{par-sum})$ is fulfilled and the condition*

$$\limsup_{n\to+\infty} \left(\frac{\lambda_n \beta_n}{\mu} + \lambda_n \beta \right) < 1$$

holds, where

$$\beta = \max\{\nu, \nu_1, ..., \nu_m\} + \sqrt{\sum_{i=1}^m \|L_i\|^2}.$$

Then $(z_n)_{n\geq 1}$ converges weakly to an element in $\operatorname{zer}\left(A+\sum_{i=1}^m L_i^* \circ (B_i \Box D_i) \circ L_i + C + N_M\right)$ as $n \to +\infty$. If, additionally, A and B_i^{-1}, $i = 1, ..., m$ are strongly monotone, then $(x_n)_{n\geq 1}$ converges strongly to the unique element in $\operatorname{zer}\left(A+\sum_{i=1}^m L_i^* \circ (B_i \Box D_i) \circ L_i + C + N_M\right)$ as $n \to +\infty$.

Proof. We start the proof by noticing that $x \in \mathcal{H}$ is a solution to Problem 4.4 if and only if there exist $v_1 \in \mathcal{G}_1, ..., v_m \in \mathcal{G}_m$ such that

$$\left\{ \begin{array}{l} 0 \in Ax + \sum_{i=1}^m L_i^* v_i + Cx + N_M(x) \\ v_i \in (B_i \Box D_i)(L_i x), i = 1, ..., m, \end{array} \right. \tag{4. 14}$$

which is nothing else than

$$\left\{ \begin{array}{l} 0 \in Ax + \sum_{i=1}^m L_i^* v_i + Cx + N_M(x) \\ 0 \in B_i^{-1} v_i + D_i^{-1} v_i - L_i x, i = 1, ..., m. \end{array} \right. \tag{4. 15}$$

In the following we endow the product space $\mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ with inner product and associated norm defined for all $(x, v_1, ..., v_m), (y, w_1, ..., w_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ as

$$\langle (x, v_1, ..., v_m), (y, w_1, ..., w_m) \rangle = \langle x, y\rangle + \sum_{i=1}^m \langle v_i, w_i \rangle$$

and

$$\|(x, v_1, ..., v_m)\| = \sqrt{\|x\|^2 + \sum_{i=1}^m \|v_i\|^2},$$

respectively.

We introduce the operators $\widetilde{A} : \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m \rightrightarrows \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$

$$\widetilde{A}(x, v_1, ..., v_m) = Ax \times B_1^{-1} v_1 \times .... \times B_m^{-1} v_m,$$

$\widetilde{D} : \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m \to \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$,

$$\widetilde{D}(x, v_1, ..., v_m) = \left(\sum_{i=1}^m L_i^* v_i + Cx, D_1^{-1} v_1 - L_1 x, ..., D_m^{-1} v_m - L_m x \right)$$

and $\widetilde{B} : \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m \to \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$,

$$\widetilde{B}(x, v_1, ..., v_m) = (Bx, 0, ..., 0).$$

Notice that, since A and B_i, $i = 1, ..., m$, are maximally monotone, $\widetilde{A}$ is maximally monotone, too (see [26, Props. 20.22, 20.23]). Further, as it was done in [76, Theorem 3.1], one can show that $\widetilde{D}$ is a monotone and β-Lipschitz continuous operator. For the sake of completeness we include here some details of the proof of these two statements.

Let be $(x, v_1, ..., v_m), (y, w_1, ..., w_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$. By using the monotonicity of C and D_i^{-1}, $i = 1, ..., m$, we have

$$\begin{aligned}
&\langle (x, v_1, ..., v_m) - (y, w_1, ..., w_m), \widetilde{D}(x, v_1, ..., v_m) - \widetilde{D}(y, w_1, ..., w_m)\rangle \\
&\qquad = \langle x - y, Cx - Cy\rangle + \sum_{i=1}^{m} \langle v_i - w_i, D_i^{-1}v_i - D_i^{-1}w_i\rangle \\
&\qquad\quad + \sum_{i=1}^{m} (\langle x - y, L_i^*(v_i - w_i)\rangle - \langle v_i - w_i, L_i(x - y)\rangle) \geq 0,
\end{aligned}$$

which shows that $\widetilde{D}$ is monotone.

The Lipschitz continuity of $\widetilde{D}$ follows by noticing that

$$\begin{aligned}
&\left\| \widetilde{D}(x, v_1, ..., v_m) - \widetilde{D}(y, w_1, ..., w_m) \right\| \\
&\leq \left\| \left(Cx - Cy, D_1^{-1}v_1 - D_1^{-1}w_1, ..., D_m^{-1}v_m - D_m^{-1}w_m\right) \right\| \\
&\quad + \left\| \left(\sum_{i=1}^{m} L_i^*(v_i - w_i), -L_1(x - y), ..., -L_m(x - y) \right) \right\| \\
&\leq \sqrt{\nu^2\|x - y\|^2 + \sum_{i=1}^{m} \nu_i^2 \|v_i - w_i\|^2} \\
&\quad + \sqrt{\left(\sum_{i=1}^{m} \|L_i\| \cdot \|v_i - w_i\| \right)^2 + \sum_{i=1}^{m} \|L_i\|^2 \cdot \|x - y\|^2} \\
&\leq \beta \|(x, v_1, ..., v_m) - (y, w_1, ..., w_m)\|.
\end{aligned}$$

Moreover, $\widetilde{B}$ is monotone, μ^{-1}-Lipschitz continuous and

$$\operatorname{zer} \widetilde{B} = \operatorname{zer} B \times \mathcal{G}_1 \times ... \times \mathcal{G}_m = M \times \mathcal{G}_1 \times ... \times \mathcal{G}_m,$$

hence

$$N_{\widetilde{M}}(x, v_1, ..., v_m) = N_M(x) \times \{0\} \times ... \times \{0\},$$

where

$$\widetilde{M} = M \times \mathcal{G}_1 \times ... \times \mathcal{G}_m = \operatorname{zer} \widetilde{B}.$$

Taking into consideration (4. 15), we obtain that $x \in \mathcal{H}$ is a solution to Problem 4.4 if and only if there exist $v_1 \in \mathcal{G}_1, ..., v_m \in \mathcal{G}_m$ such that

$$(x, v_1, ..., v_m) \in \operatorname{zer}(\widetilde{A} + \widetilde{D} + N_{\widetilde{M}}).$$

Conversely, when $(x, v_1, ..., v_m) \in \operatorname{zer}(\widetilde{A} + \widetilde{D} + N_{\widetilde{M}})$, then $x \in \operatorname{zer}\big(A + \sum_{i=1}^{m} L_i^* \circ (B_i \Box D_i) \circ L_i + C + N_M\big)$. This means that determining the zeros of $\widetilde{A} + \widetilde{D} + N_{\widetilde{M}}$ will automatically provide a solution to Problem 4.4.

Using that

$$J_{\lambda\widetilde{A}}(x, v_1, ..., v_m) = (J_{\lambda A}(x), J_{\lambda B_1^{-1}}(v_1), ..., J_{\lambda B_m^{-1}}(v_m))$$

for every $(x, v_1, ..., v_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ and every $\lambda > 0$ (see [26, Proposition 23.16]), one can easily see that the iterations of Algorithm 4.4 read for all $n \geq 1$:

$$\left\{\begin{array}{l} (p_n, q_{1,n}, ..., q_{m,n}) = J_{\lambda_n\widetilde{A}}\Big[(x_n, v_{1,n}, ..., v_{m,n}) - \lambda_n\widetilde{D}(x_n, v_{1,n}, ..., v_{m,n}) \\ \qquad\qquad -\lambda_n\beta_n\widetilde{B}(x_n, v_{1,n}, ..., v_{m,n})\Big] \\ (x_{n+1}, v_{1,n+1}, ..., v_{m,n+1}) = \lambda_n\beta_n\Big[\widetilde{B}(x_n, v_{1,n}, ..., v_{m,n}) - \widetilde{B}(p_n, q_{1,n}, ..., q_{m,n})\Big] \\ \qquad\qquad +\lambda_n\Big[\widetilde{D}(x_n, v_{1,n}, ..., v_{m,n}) - \widetilde{D}(p_n, q_{1,n}, ..., q_{m,n})\Big] \\ \qquad\qquad +(p_n, q_{1,n}, ..., q_{m,n}), \end{array}\right.$$

which is nothing else than the iterative scheme of Algorithm 4.3 employed to the monotone inclusion problem

$$0 \in \widetilde{A}x + \widetilde{D}x + N_{\widetilde{M}}(x).$$

In order to compute the Fitzpatrick function of $\widetilde{B}$, we consider arbitrary elements $(x, v_1, ..., v_m), (x', v_1', ..., v_m') \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$. It holds

$$\begin{aligned} &\varphi_{\widetilde{B}}\left((x, v_1, ..., v_m), (x', v_1', ..., v_m')\right) \\ &= \sup_{\substack{(y,w_1,...,w_m)\in \\ \mathcal{H}\times\mathcal{G}_1\times...\times\mathcal{G}_m}} \Big\{\langle (x, v_1, ..., v_m), \widetilde{B}(y, w_1, ..., w_m)\rangle \\ &\quad + \langle (x', v_1', ..., v_m'), (y, w_1, ..., w_m)\rangle - \langle (y, w_1, ..., w_m), \widetilde{B}(y, w_1, ..., w_m)\rangle\Big\} \\ &= \sup_{\substack{(y,w_1,...,w_m)\in \\ \mathcal{H}\times\mathcal{G}_1\times...\times\mathcal{G}_m}} \left\{\langle x, By\rangle + \langle x', y\rangle + \sum_{i=1}^m \langle v_i', w_i\rangle - \langle y, By\rangle\right\}, \end{aligned}$$

thus

$$\varphi_{\widetilde{B}}\big((x, v_1, ..., v_m), (x', v_1', ..., v_m')\big) = \left\{\begin{array}{ll} \varphi_B(x, x'), & \text{if } v_1' = ... = v_m' = 0, \\ +\infty, & \text{otherwise.} \end{array}\right.$$

Moreover,

$$\sigma_{\widetilde{M}}(x, v_1, ..., v_m) = \left\{\begin{array}{ll} \sigma_M(x), & \text{if } v_1 = ... = v_m = 0, \\ +\infty, & \text{otherwise,} \end{array}\right.$$

hence condition (ii) in $(H_{fitz}^{par-sum})$ is nothing else than: for each $(p, p_1, ..., p_m) \in \operatorname{ran} N_{\widetilde{M}} = \operatorname{ran} N_M \times \{0\} \times ... \times \{0\}$ one has

$$\begin{aligned} &\sum_{n\geq 1} \lambda_n\beta_n\Big[\sup_{(u,v_1,...,v_m)\in\widetilde{M}} \varphi_{\widetilde{B}}\Big((u, v_1, ..., v_m), \frac{(p, p_1, ..., p_m)}{\beta_n}\Big) \\ &- \sigma_{\widetilde{M}}\Big(\frac{(p, p_1, ..., p_m)}{\beta_n}\Big)\Big] < +\infty. \end{aligned}$$

Furthermore, condition (i) in $(H_{fitz}^{par-sum})$ ensures that $\widetilde{A}+N_{\widetilde{M}}$ is maximally monotone and $\operatorname{zer}(\widetilde{A}+\widetilde{D}+N_{\widetilde{M}}) \neq \emptyset$. Hence, we are in the position of applying Theorem 4.4 in the context of finding the zeros of $\widetilde{A}+\widetilde{D}+N_{\widetilde{M}}$. The statements of the theorem are an easy consequence of this result. □

4.2.3 Convex minimization problems

In this subsection we employ the results given for monotone inclusions in the special instance when minimizing a convex function with an intricate formulation with respect to the set of minimizers of another convex and differentiable function with Lipschitz continuous gradient.

Problem 4.5 *Let $\mathcal{H}$ be a real Hilbert space, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with a ν-Lipschitz continuous gradient for $\nu > 0$. Let m be a strictly positive integer and for every $i = 1, ..., m$ let $\mathcal{G}_i$ be a real Hilbert space, $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that l_i is ν_i^{-1}-strongly convex for $\nu_i > 0$ and $L_i : \mathcal{H} \to \mathcal{G}_i$ a nonzero linear continuous operator. Further, let $\Psi \in \Gamma(\mathcal{H})$ be differentiable with a μ^{-1}-Lipschitz continuous gradient, fulfilling $\min \Psi = 0$. The convex minimization problem under investigation is*

$$\inf_{x \in \operatorname{argmin} \Psi} \left\{ f(x) + \sum_{i=1}^{m} (g_i \Box l_i)(L_i x) + h(x) \right\}. \tag{4. 16}$$

Consider the maximal monotone operators

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

According to [26, Proposition 17.10, Theorem 18.15], $D_i^{-1} = \nabla l_i^*$ is a monotone and ν_i-Lipschitz continuous operator for $i = 1, ..., m$. Moreover, B is a monotone and μ^{-1}-Lipschitz continuous operator and

$$M := \operatorname{argmin} \Psi = \operatorname{zer} B.$$

Taking into account the sum rules of the convex subdifferential, every element of $\operatorname{zer}\left(\partial f + \sum_{i=1}^m L_i^* \circ (\partial g_i \Box \partial l_i) \circ L_i + \nabla h + N_M\right)$ is an optimal solution of (4. 16). The converse is true if an appropriate qualification condition is satisfied. For the readers convenience, we present some qualification conditions which are suitable in this context. One of the weakest qualification conditions of interiority-type reads (see, for instance, [76, Proposition 4.3, Remark 4.4])

$$(0, ..., 0) \in \operatorname{sqri} \left(\prod_{i=1}^{m} (\operatorname{dom} g_i + \operatorname{dom} l_i) - \{(L_1 x, ..., L_m x) : x \in \operatorname{dom} f \cap M\} \right). \tag{4. 17}$$

The condition (4. 17) is fulfilled if one of the following conditions holds (see for example [76, Proposition 4.3]):

(i) $\operatorname{dom} g_i + \operatorname{dom} l_i = \mathcal{G}_i$, $i = 1, ..., m$;

(ii) $\mathcal{H}$ and $\mathcal{G}_i$ are finite-dimensional and there exists $x \in \operatorname{ri} \operatorname{dom} f \cap \operatorname{ri} M$ such that $L_i x \in \operatorname{ri} \operatorname{dom} g_i + \operatorname{ri} \operatorname{dom} l_i$, $i = 1, ..., m$.

Algorithm 4.4 becomes in this particular case

Algorithm 4.5

`Initialization:` `Choose` $(x_1, v_{1,1}, ..., v_{m,1}) \in \mathcal{H} \times \mathcal{G}_1 \times ... \times \mathcal{G}_m$

`For` $n \geq 1$ `set:`

$$\begin{aligned}
p_n &= \operatorname{prox}_{\lambda_n f}[x_n - \lambda_n(\nabla h(x_n) + \textstyle\sum_{i=1}^m L_i^* v_{i,n}) - \lambda_n \beta_n \nabla \Psi(x_n)] \\
q_{i,n} &= \operatorname{prox}_{\lambda_n g_i^*}[v_{i,n} + \lambda_n(L_i x_n - \nabla l_i^*(v_{i,n}))],\ i = 1, ..., m \\
x_{n+1} &= \lambda_n \beta_n (\nabla \Psi(x_n) - \nabla \Psi(p_n)) + \lambda_n(\nabla h(x_n) - \nabla h(p_n)) \\
&\quad + \lambda_n \textstyle\sum_{i=1}^m L_i^*(v_{i,n} - q_{i,n}) + p_n \\
v_{i,n+1} &= \lambda_n L_i(p_n - x_n) + \lambda_n(\nabla l_i^*(v_{i,n}) - \nabla l_i^*(q_{i,n})) \\
&\quad + q_{i,n},\ i = 1, ..., m
\end{aligned}$$

For the convergence result we need the following hypotheses:

$$(H_{fitz}^{opt}) \begin{cases} (i)\ \partial f + N_M \text{ is maximally monotone and} \\ \quad\ \ (4.\ 16) \text{ has an optimal solution;} \\ (ii)\ \ \text{For every } p \in \operatorname{ran} N_M, \sum_{n\geq 1} \lambda_n\beta_n \left[\Psi^*\left(\frac{p}{\beta_n}\right) - \sigma_M\left(\frac{p}{\beta_n}\right)\right] < +\infty; \\ (iii)\ (\lambda_n)_{n\geq 1} \in \ell^2 \setminus \ell^1. \end{cases}$$

Remark 4.7 (a) Let us mention that $\partial f + N_M$ is maximally monotone, if

$$0 \in \operatorname{sqri}(\operatorname{dom} f - M),$$

a condition which is fulfilled if, for instance,

$$f \text{ is continuous at a point in } \operatorname{dom} f \cap M$$

or

$$\operatorname{int} M \cap \operatorname{dom} f \neq \emptyset.$$

(b) Since $\Psi(x) = 0$ for all $x \in M$, by (4. 1) it follows that whenever (ii) in (H_{fitz}^{opt}) holds, condition (ii) in $(H_{fitz}^{par-sum})$, formulated for $B = \nabla\Psi$, is also true.

(c) Let us mention that hypothesis (ii) is satisfied, if

$$\sum_{n\geq 1} \frac{\lambda_n}{\beta_n} < +\infty$$

and Ψ is bounded below by a multiple of the square of the distance to C (see [15]). This is for instance the case when

$$M = \operatorname{zer} L,$$

where $L : \mathcal{H} \to \mathcal{H}$ is a linear continuous operator with closed range and

$$\Psi : \mathcal{H} \to \mathbb{R}, \Psi(x) = \|Lx\|^2$$

(see [15, 16]). For further situations for which condition (ii) is fulfilled we refer to [16, Section 4.1] (see also [24]).

We are able now to formulate the convergence result.

Theorem 4.7 *Consider the sequences generated by Algorithm 4.5 and $(z_n)_{n\geq 1}$ the sequence defined in (4. 2). If (H_{fitz}^{opt}) and (4. 17) are fulfilled and*

$$\limsup_{n\to+\infty} \left(\frac{\lambda_n\beta_n}{\mu} + \lambda_n\beta\right) < 1,$$

where

$$\beta = \max\{\nu, \nu_1, ..., \nu_m\} + \sqrt{\sum_{i=1}^m \|L_i\|^2},$$

then $(z_n)_{n\geq 1}$ converges weakly to an optimal solution to (4. 16) as $n \to +\infty$. If, additionally, f and g_i^*, $i = 1, ..., m$ are strongly convex, then $(x_n)_{n\geq 1}$ converges strongly to the unique optimal solution of (4. 16) as $n \to +\infty$.

Remark 4.8 (a) According to [26, Proposition 17.10, Theorem 18.15], for a function $g \in \Gamma(\mathcal{H})$ one has that g is strongly convex if and only if g is differentiable with Lipschitz continuous gradient.

(b) Notice that in case $\Psi(x) = 0$ for all $x \in \mathcal{H}$, Algorithm 4.5 turns out to be the error-free variant of the iterative scheme given in [76, Theorem 4.2] for solving the convex minimization problem

$$\inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^m (g_i \Box l_i)(L_i x) + h(x) \right\}. \tag{4. 18}$$

4.2.4 A numerical experiment in TV-based image inpainting

In this section we illustrate the applicability of Algorithm 4.5 when solving an image inpainting problem, which aims for recovering lost information. We consider images of size $M \times N$ as vectors $x \in \mathbb{R}^n$ for $n = M \cdot N$, while each pixel denoted by $x_{i,j}$, $1 \leq i \leq M$, $1 \leq j \leq N$, ranges in the closed interval from 0 (pure black) to 1 (pure white). We denote by $b \in \mathbb{R}^n$ the image with missing pixels (in our case set to black) and by $K \in \mathbb{R}^{n \times n}$ the diagonal matrix with $K_{i,i} = 0$, if the pixel i in the noisy image $b \in \mathbb{R}^n$ is missing, and $K_{i,i} = 1$, otherwise, $i = 1, ..., n$ (notice that $\|K\| = 1$). The original image will be reconstructed by considering the following TV-regularized model

$$\inf \left\{ TV_{\text{iso}}(x) : Kx = b, x \in [0,1]^n \right\}. \tag{4. 19}$$

The objective function $TV_{\text{iso}} : \mathbb{R}^n \to \mathbb{R}$ is the isotropic total variation and we refer the reader to the section concerning numerical experiments in Chapter 2 for its definition. By using also the notations introduced there, and by considering the function $\Psi : \mathbb{R}^n \to \mathbb{R}$,

$$\Psi(x) = \frac{1}{2}\|Kx - b\|^2,$$

problem (4. 19) can be reformulated as

$$\inf_{x \in \operatorname{argmin} \Psi} \left\{ f(x) + g_1(Lx) \right\}, \tag{4. 20}$$

where $f : \mathbb{R}^n \to \mathbb{R}$,

$$f = \delta_{[0,1]^n}$$

and $g_1 : \mathcal{Y} \to \mathbb{R}$,

$$g_1(y_1, y_2) = \|(y_1, y_2)\|_\times.$$

Problem (4. 20) is of type (4. 16), when one takes $m = 1$, $L_1 = L$, $l_1 = \delta_{\{0\}}$ and $h = 0$. Notice that $\nabla\Psi(x) = K(Kx - b) = K(x - b)$ for every $x \in \mathbb{R}^n$, thus $\nabla\Psi$ is Lipschitz continuous with Lipschitz constant $\mu = 1$. The iterative scheme in Algorithm 4.5 becomes for every $n \geq 0$ in this case

$$\left\{\begin{array}{l} p_n = \operatorname{prox}_{\lambda_n f}[x_n - \lambda_n L^* v_{1,n} - \lambda_n \beta_n K(x_n - b)] \\ q_{1,n} = \operatorname{prox}_{\lambda_n g_1^*}(v_{1,n} + \lambda_n L x_n) \\ x_{n+1} = \lambda_n \beta_n K(x_n - p_n) + \lambda_n L^*(v_{1,n} - q_{1,n}) + p_n \\ v_{1,n+1} = \lambda_n L(p_n - x_n) + q_{1,n}. \end{array}\right.$$

For the proximal points we have the following formulae:

$$\operatorname{prox}_{\gamma f}(x) = \operatorname{proj}_{[0,1]^n}(x) \ \forall \gamma > 0 \text{ and } \forall x \in \mathbb{R}^n$$

and (see [58])

$$\operatorname{prox}_{\gamma g_1^*}(p, q) = \operatorname{proj}_S(p, q) \ \ \forall \gamma > 0 \text{ and } \forall (p, q) \in \mathcal{Y},$$

where

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

and the projection operator $\operatorname{proj}_S : \mathcal{Y} \to S$ is defined via

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

We tested the algorithm on the fruit image and considered as parameters $\lambda_n = 0.9 \cdot n^{-0.75}$ and $\beta_n = n^{0.75}$ for all $n \geq 1$. Figure 4.1 shows the original image, the image obtained from it after setting 80% randomly chosen pixels to black, the nonaveraged reconstructed image x^n and the averaged reconstructed image z^n after 1000 iterations.

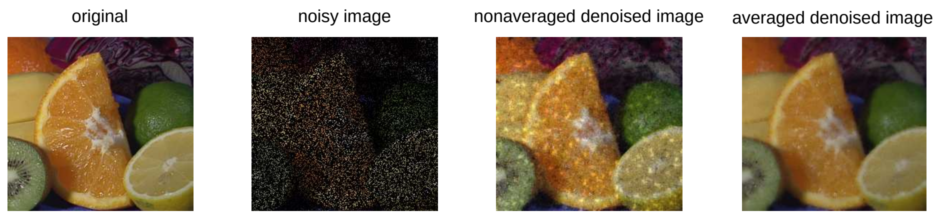


Figure 4.1: TV image inpainting: the original image, the image with 80% missing pixels, the nonaveraged reconstructed image x^n and the averaged reconstructed image z^n after 1000 iterations.

The comparisons concerning the quality of the reconstructed images were made by means of the improvement in signal-to-noise ratio (ISNR), which is defined as

$$\mathrm{ISNR}(n) = 10 \log_{10} \left(\frac{\|x - b\|^2}{\|x - x^n\|^2} \right),$$

where x, b and x^n denote the original, the image with missing pixels and the recovered image at iteration n, respectively.

Figure 4.2 shows the evolution of the ISNR values for the averaged and the nonaveraged reconstructed images. Both figures illustrate the theoretical outcomes concerning the sequences involved in Theorem 4.7, namely that the averaged sequence has better convergence properties than the nonaveraged one.

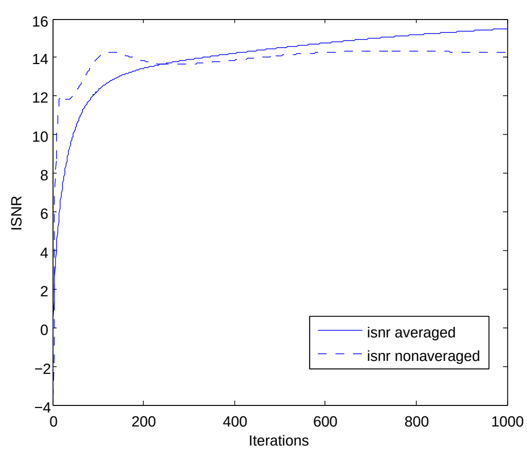


Figure 4.2: The ISNR curves for the averaged and nonaveraged reconstructed images

Chapter 5

Implicit-type dynamical systems associated with monotone inclusion problems

In this chapter we approach the solving of monotone inclusion problems of the form (1. 3) by investigating dynamical systems of implicit-type formulated via the resolvents of the operators involved. In Section 5.1 we consider first-order dynamical systems and investigate the asymptotic properties of the trajectories, obtaining also convergence rates and in Section 5.2 we focus our attention on second order dynamics.

We recall first some technical results and specific tools which will be used in this framework. We consider the following definition of an absolutely continuous function, see also [2, 20].

Definition 5.1 (see, for instance, [2,20]) A function $x : [0, b] \to \mathcal{H}$ (where $b > 0$) is said to be absolutely continuous if one of the following equivalent properties holds:

(i) there exists an integrable function $y : [0, b] \to \mathcal{H}$ such that

$$x(t) = x(0) + \int_0^t y(s)ds \quad \forall t \in [0, b];$$

(ii) x is continuous and its distributional derivative is Lebesgue integrable on $[0, b]$;

(iii) for every $\varepsilon > 0$, there exists $\eta > 0$ such that for any finite family of intervals $I_k = (a_k, b_k) \subseteq [0, b]$ we have the implication

$$\left(I_k \cap I_j = \emptyset \text{ and } \sum_k |b_k - a_k| < \eta \right) \Longrightarrow \sum_k \|x(b_k) - x(a_k)\| < \varepsilon.$$

Remark 5.1 (a) It follows from the definition that an absolutely continuous function is differentiable almost everywhere, its derivative coincides with its distributional derivative almost everywhere and one can recover the function from its derivative $\dot{x} = y$ by the integration formula (i).

(b) If $x : [0, b] \to \mathcal{H}$, where $b > 0$, is absolutely continuous and $B : \mathcal{H} \to \mathcal{H}$ is L-Lipschitz continuous for $L \geq 0$, then the function $z = B \circ x$ is absolutely continuous, too. This can be easily seen by using the characterization of absolute continuity in Definition 5.1(iii). Moreover, z is differentiable almost everywhere on $[0, b]$ and the inequality $\|\dot{z}(t)\| \leq L\|\dot{x}(t)\|$ holds for almost every $t \in [0, b]$.

The following two well-known results, which can be interpreted as continuous versions of the quasi-Fejér monotonicity for sequences, will play an important role in the asymptotic analysis of the trajectories of several dynamical systems investigated in this chapter. For their proofs we refer the reader to [2, Lemma 5.1] and [2, Lemma 5.2], respectively.

Lemma 5.1 *Suppose that $F : [0, +\infty) \to \mathbb{R}$ is locally absolutely continuous and bounded from below and that there exists $G \in L^1([0, +\infty))$ such that for almost every $t \in [0, +\infty)$*

$$\frac{d}{dt}F(t) \le G(t).$$

Then there exists $\lim_{t\to\infty} F(t) \in \mathbb{R}$.

Lemma 5.2 *If $1 \le p < \infty$, $1 \le r \le \infty$, $F : [0, +\infty) \to [0, +\infty)$ is locally absolutely continuous, $F \in L^p([0, +\infty))$, $G : [0, +\infty) \to \mathbb{R}$, $G \in L^r([0, +\infty))$ and for almost every $t \in [0, +\infty)$*

$$\frac{d}{dt}F(t) \le G(t),$$

then $\lim_{t\to+\infty} F(t) = 0$.

The next result which we recall here is the continuous version of the Opial Lemma (see, for example, [2, Lemma 5.3], [1, Lemma 2.10]).

Lemma 5.3 *Let $S \subseteq \mathcal{H}$ be a nonempty set and $x : [0, +\infty) \to \mathcal{H}$ a given map. Assume that*

(i) for every $x^* \in S$, $\lim_{t\to+\infty} \|x(t) - x^*\|$ exists;

(ii) every weak sequential cluster point of the map x belongs to S.

Then there exists $x_\infty \in S$ such that $x(t)$ converges weakly to x_∞ as $t \to +\infty$.

5.1 First order dynamical systems

This section is dedicated to the asymptotic analysis of the trajectories of first order dynamical systems associated to monotone inclusion problems.

5.1.1 First order dynamical systems for monotone inclusion problems

We start with studying a dynamical systems associated to the fixed points set of a nonexpansive operator. Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping $\lambda : [0, +\infty) \to [0, 1]$ be a Lebesgue measurable function and $x_0 \in \mathcal{H}$. For the beginning we are concerned with the following dynamical system:

$$\left\{ \begin{array}{l} \dot{x}(t) = \lambda(t)\big(T(x(t)) - x(t)\big) \\ x(0) = x_0. \end{array} \right. \tag{5. 1}$$

The first issue we investigate is the existence of strong solutions for (5. 1).

Definition 5.2 We say that $x : [0, +\infty) \to \mathcal{H}$ is a strong global solution of (5. 1) if the following properties are satisfied:

(i) $x : [0, +\infty) \to \mathcal{H}$ is absolutely continuous on each interval $[0, b]$, $0 < b < +\infty$;

(ii) $\dot{x}(t) = \lambda(t)\big(T(x(t)) - x(t)\big)$ for almost every $t \in [0, +\infty)$;

(iii) $x(0) = x_0$.

In what follows we verify the existence and uniqueness of strong global solutions of (5. 1). To this end we use the Cauchy-Lipschitz theorem for absolutely continues trajectories (see for example [90, Proposition 6.2.1], [125, Theorem 54]).

It is immediate that the system (5. 1) can be written as

$$\begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(0) = x_0, \end{cases} \tag{5. 2}$$

where $f : [0, +\infty) \times \mathcal{H} \to \mathcal{H}$ is defined by

$$f(t, x) = \lambda(t)(Tx - x).$$

(a) Take arbitrary $x, y \in \mathcal{H}$. Relying on the nonexpansiveness of T, for all $t \geq 0$ we have

$$\|f(t, x) - f(t, y)\| \leq 2\lambda(t)\|x - y\|.$$

Since λ is bounded above, one has $2\lambda(\cdot) \in L^1([0, b])$ for any $0 < b < +\infty$;

(b) Take arbitrary $x \in \mathcal{H}$ and $b > 0$. One has

$$\int_0^b \|f(t, x)\|dt = \|Tx - x\| \int_0^b \lambda(t)dt \leq b\|Tx - x\|,$$

hence

$$\forall x \in \mathcal{H},\ \forall b > 0, \quad f(\cdot, x) \in L^1([0, b], \mathcal{H}).$$

By considering the statements proven in (a) and (b), the existence and uniqueness of a strong global solution of the dynamic system (5. 1) follows.

Remark 5.2 (i) From the considerations above one can easily notice that the existence and uniqueness of strong global solutions of (5. 1) can be guaranteed in the more general setting when T is Lipschitz continuous and $\lambda : [0, +\infty) \to \mathbb{R}$ is a Lebesgue measurable function such that $\lambda(\cdot) \in L^1_{\text{loc}}([0, +\infty))$.

(ii) Let us mention that if we suppose additionally that λ is continuous, then the global version of the Picard-Lindelöf Theorem allows us to conclude that, for $x_0 \in \mathcal{H}$, there exists a unique trajectory $x : [0, +\infty) \to \mathcal{H}$ which is a C^1 function and which satisfies the relation (ii) in Definition 5.2 for every $t \in [0, +\infty)$.

In the following we investigate the convergence properties of the trajectories of the dynamical system (5. 1). We show that under mild conditions imposed on the function λ, the orbits converge weakly to a fixed point of the nonexpansive operator, provided the set of such points is nonempty.

Theorem 5.1 *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping such that $\operatorname{Fix} T \neq \emptyset$, $\lambda : [0, +\infty) \to [0, 1]$ a Lebesgue measurable function and $x_0 \in \mathcal{H}$. Suppose that one of the following conditions is fulfilled:*

$$\int_0^{+\infty} \lambda(t)(1 - \lambda(t))dt = +\infty \ \textit{or} \ \inf_{t \geq 0} \lambda(t) > 0.$$

Let $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of (5. 1). Then the following statements are true:

(i) the trajectory x is bounded and $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$;

(ii) $\lim_{t \to +\infty}(T(x(t)) - x(t)) = 0$;

(iii) $\lim_{t \to +\infty} \dot{x}(t) = 0$;

(iv) $x(t)$ converges weakly to a point in $\operatorname{Fix} T$, as $t \to +\infty$.

Proof. We rely on Lyapunov analysis combined with the Opial Lemma. We take an arbitrary $x^* \in \operatorname{Fix} T$ and give an estimation for $\frac{d}{dt}\|x(t)-x^*\|^2$. By (1. 28), the fact that $x^* \in \operatorname{Fix} T$ and the nonexpansiveness of T we obtain for almost every $t \geq 0$:

$$\begin{aligned}
\frac{d}{dt}\|x(t)-x^*\|^2 =& 2\langle \dot{x}(t), x(t)-x^*\rangle = \|\dot{x}(t)+x(t)-x^*\|^2 - \|x(t)-x^*\|^2 - \|\dot{x}(t)\|^2\\
=&\|\lambda(t)(T(x(t))-x^*)+(1-\lambda(t))(x(t))-x^*)\|^2\\
&-\|x(t)-x^*\|^2-\|\dot{x}(t)\|^2\\
=&\lambda(t)\|T(x(t))-x^*\|^2+(1-\lambda(t))\|x(t)-x^*\|^2\\
&-\lambda(t)(1-\lambda(t))\|T(x(t)-x(t))\|^2-\|x(t)-x^*\|^2-\|\dot{x}(t)\|^2\\
\leq&-\lambda(t)(1-\lambda(t))\|T(x(t)-x(t))\|^2-\|\dot{x}(t)\|^2.
\end{aligned}$$

Hence for almost every $t \geq 0$ we have that

$$\frac{d}{dt}\|x(t)-x^*\|^2+\lambda(t)(1-\lambda(t))\|T(x(t)-x(t))\|^2+\|\dot{x}(t)\|^2 \leq 0. \tag{5. 3}$$

Since $\lambda(t) \in [0,1]$ for all $t \geq 0$, from (5. 3) it follows that $t \mapsto \|x(t)-x^*\|$ is decreasing, hence $\lim_{t\to+\infty}\|x(t)-x^*\|$ exists. From here we obtain the boundedness of the trajectory and by integrating (5. 3) we deduce also that

$$\int_0^{+\infty}\|\dot{x}(t)\|^2 dt < +\infty$$

and

$$\int_0^{+\infty}\lambda(t)(1-\lambda(t))\|T(x(t))-x(t)\|^2 dt < +\infty, \tag{5. 4}$$

thus (i) holds. Since $x^* \in \operatorname{Fix} T$ has been chosen arbitrary, the first assumption in the continuous version of Opial Lemma is fulfilled.

We show in the following that $\lim_{t\to+\infty}(T(x(t))-x(t))$ exists and it is a real number. This is immediate if we show that the function $t \mapsto \frac{1}{2}\|T(x(t))-x(t)\|^2$ is decreasing. According to Remark 5.1(b), the function $t \mapsto T(x(t))$ is almost everywhere differentiable and $\|\frac{d}{dt}T(x(t))\| \leq \|\dot{x}(t)\|$ holds for almost every $t \geq 0$. Moreover, by the first equation of (5. 1) we have

$$\begin{aligned}
\frac{d}{dt}\left(\frac{1}{2}\|T(x(t))-x(t)\|^2\right) =& \left\langle \frac{d}{dt}T(x(t))-\dot{x}(t), T(x(t))-x(t)\right\rangle\\
=&-\langle \dot{x}(t), T(x(t))-x(t)\rangle + \left\langle \frac{d}{dt}T(x(t)), T(x(t))-x(t)\right\rangle\\
=&-\lambda(t)\|T(x(t))-x(t)\|^2+\left\langle \frac{d}{dt}T(x(t)), T(x(t))-x(t)\right\rangle\\
\leq&-\lambda(t)\|T(x(t))-x(t)\|^2+\|\dot{x}(t)\|\cdot\|T(x(t))-x(t)\| = 0,
\end{aligned}$$

hence $\lim_{t\to+\infty}(T(x(t))-x(t))$ exists and is a real number.

(a) Firstly, let us assume that $\int_0^{+\infty}\lambda(t)(1-\lambda(t))dt = +\infty$. This immediately implies by (5. 4) that $\lim_{t\to+\infty}(T(x(t))-x(t)) = 0$, thus (ii) holds. Taking into account that λ is bounded, from (5. 1) and (ii) we deduce (iii). For the last property of the theorem we need to verify the second assumption of the Opial Lemma. Let $\overline{x} \in \mathcal{H}$ be a weak sequential cluster point of x, that is, there exists a sequence $t_n \to +\infty$ (as $n \to \infty$) such that $(x(t_n))_{n\in\mathbb{N}}$ converges weakly to $\overline{x}$. Applying Lemma 1.3 and (ii) we obtain $\overline{x} \in \operatorname{Fix} T$ and the conclusion follows.

(b) We suppose now that $\inf_{t\geq 0}\lambda(t) > 0$. From the first relation of (5. 1) and (i) we easily deduce that $Tx - x \in L^2([0,\infty),\mathcal{H})$, hence the function $t \mapsto \frac{1}{2}\|T(x(t))-$

$x(t)\|^2$ belongs to $L^1([0,\infty))$. Since $\frac{d}{dt}\left(\frac{1}{2}\|T(x(t))-x(t)\|^2\right) \leq 0$ for almost every $t \geq 0$, we obtain by applying Lemma 5.2 that $\lim_{t\to\infty} \|T(x(t))-x(t)\|^2 = 0$, thus (ii) holds. The rest of the proof can be done in the lines of case (a) considered above. □

Remark 5.3 Let us specify that due to the fact that the equality in Definition 5.2(ii) holds almost everywhere, the conclusion in Theorem 5.1(iii) (which has been obtained as a consequence of Theorem 5.1(ii)) has to be considered in the *almost-limit* sense (see also [22, Definition 1, Chapter 6, Section 5]), which means that in the classical definition of the limit, the required inequality holds almost everywhere.

Remark 5.4 Notice that the function $\lambda_1(t) = \frac{1}{t+1}$, for all $t \geq 0$, verifies the condition $\int_0^{+\infty} \lambda_1(t)(1-\lambda_1(t))dt = +\infty$, while $\inf_{t\geq 0} \lambda_1(t) > 0$ is not fulfilled. On the other hand, the function $\lambda_2(t) = 1$, for all $t \geq 0$, verifies the condition $\inf_{t\geq 0} \lambda_2(t) > 0$, while $\int_0^{\infty} \lambda_2(t)(1-\lambda_2(t))dt = \infty$ fails. This shows that the two assumptions on λ under which the conclusions of Theorem (5.1) are valid are independent.

Remark 5.5 The explicit discretization of (5. 1) with respect to the time variable t, with step size $h_n > 0$, yields for an initial point x_0 the following iterative scheme:

$$x_{n+1} = x_n + h_n\lambda_n(Tx_n - x_n) \ \forall n \geq 0.$$

By taking $h_n = 1$ this becomes

$$x_{n+1} = x_n + \lambda_n(Tx_n - x_n) \ \forall n \geq 0, \tag{5. 5}$$

which is the classical Krasnosel'skiĭ–Mann algorithm for finding the set of fixed points of the nonexpansive operator T (see [26, Theorem 5.14]). The convergence of (5. 5) is guaranteed under the condition

$$\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty.$$

Notice that in case $\lambda_n = 1$ for all $n \in \mathbb{N}$ and for an initial point x_0 different from 0, the convergence of (5. 5) can fail, as it happens for instance for the operator $T = -\operatorname{Id}$. In contrast to this, as pointed out in Theorem 5.1, the dynamical system (5. 1) has a strong global solution and the convergence of the trajectory is guaranteed also in case $\lambda(t) = 1$ for all $t \geq 0$.

An immediate consequence of Theorem 5.1 is the following corollary, where we consider dynamical systems involving averaged operators.

Corollary 5.1 *Let $\alpha \in (0,1)$, $R : \mathcal{H} \to \mathcal{H}$ be α-averaged such that* $\operatorname{Fix} R \neq \emptyset$, *$\lambda : [0,+\infty) \to [0, 1/\alpha]$ a Lebesgue measurable function and $x_0 \in \mathcal{H}$. Suppose that one of the following conditions is fulfilled:*

$$\int_0^{+\infty} \lambda(t)(1-\alpha\lambda(t))dt = +\infty \ \textit{or} \ \inf_{t\geq 0} \lambda(t) > 0.$$

Let $x : [0,+\infty) \to \mathcal{H}$ be the unique strong global solution of the dynamical system

$$\begin{cases} \dot{x}(t) = \lambda(t)\big(R(x(t)) - x(t)\big) \\ x(0) = x_0. \end{cases} \tag{5. 6}$$

Then the following statements are true:

(i) the trajectory x is bounded and $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$;

(ii) $\lim_{t\to+\infty}(R(x(t)) - x(t)) = 0$;

(iii) $\lim_{t\to+\infty} \dot{x}(t) = 0$;

(iv) $x(t)$ *converges weakly to a point in* $\operatorname{Fix} R$, *as* $t \to +\infty$.

Proof. Since R is α-averaged, there exists a nonexpansive operator $T : \mathcal{H} \to \mathcal{H}$ such that $R = (1-\alpha)\operatorname{Id} + \alpha T$. The conclusion follows by taking into account that (5. 6) is equivalent to

$$\begin{cases} \dot{x}(t) = \alpha\lambda(t)\big(T(x(t)) - x(t)\big) \\ x(0) = x_0 \end{cases}$$

and $\operatorname{Fix} R = \operatorname{Fix} T$. □

In the following we investigate the convergence rate of the trajectories of the dynamical system (5. 1). This will be done in terms of the fixed point residual function $t \mapsto \|Tx(t) - x(t)\|$ and of $t \mapsto \|\dot{x}(t)\|$. Notice that convergence rates for the discrete iteratively generated algorithm (5. 5) have been investigated in [80, 82, 95].

Theorem 5.2 *Let* $T : \mathcal{H} \to \mathcal{H}$ *be a nonexpansive mapping such that* $\operatorname{Fix} T \neq \emptyset$, $\lambda : [0, +\infty) \to [0, 1]$ *a Lebesgue measurable function and* $x_0 \in \mathcal{H}$. *Suppose that*

$$0 < \inf_{t\geq 0} \lambda(t) \leq \sup_{t\geq 0} \lambda(t) < 1.$$

Let $x : [0, +\infty) \to \mathcal{H}$ *be the unique strong global solution of* (5. 1). *Then for almost every* $t \geq 0$ *we have*

$$\|\dot{x}(t)\| \leq \|T(x(t)) - x(t)\| \leq \frac{d(x_0, \operatorname{Fix} T)}{\sqrt{\underline{\tau} t}},$$

where $\underline{\tau} = \inf_{t\geq 0} \lambda(t)(1 - \lambda(t)) > 0$.

Proof. Take an arbitrary $x^* \in \operatorname{Fix} T$ and $t > 0$. From (5. 3) we have for almost every $s \geq 0$:

$$\frac{d}{ds}\|x(s) - x^*\|^2 + \lambda(s)(1 - \lambda(s))\|T(x(s) - x(s))\|^2 \leq 0. \tag{5. 7}$$

By integrating we obtain

$$\int_0^t \lambda(s)(1 - \lambda(s))\|T(x(s)) - x(s)\|^2 ds \leq \|x_0 - x^*\|^2 - \|x(t) - x^*\|^2 \leq \|x_0 - x^*\|^2.$$

We have seen in the proof of Theorem 5.1 that $t \mapsto \frac{1}{2}\|T(x(t)) - x(t)\|^2$ is decreasing, thus the last inequality yields

$$t\underline{\tau}\|T(x(t)) - x(t)\|^2 \leq \|x_0 - x^*\|^2.$$

Since this inequality holds for an arbitrary $x^* \in \operatorname{Fix} T$, we get for all $t \geq 0$:

$$\sqrt{t\underline{\tau}}\|T(x(t)) - x(t)\| \leq d(x_0, \operatorname{Fix} T).$$

By taking also into account (5. 1), the conclusion follows. □

Next we show that the convergence rates of the fixed point residual function $t \mapsto \|Tx(t) - x(t)\|$ and of $t \mapsto \|\dot{x}(t)\|$ can be improved to $o\left(\frac{1}{\sqrt{t}}\right)$.

Theorem 5.3 *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping such that* $\operatorname{Fix} T \neq \emptyset$, $\lambda : [0, +\infty) \to [0, 1]$ *a Lebesgue measurable function and $x_0 \in \mathcal{H}$. Suppose that*

$$0 < \inf_{t \geq 0} \lambda(t) \leq \sup_{t \geq 0} \lambda(t) < 1.$$

Let $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of (5. 1)*. Then for almost every $t \geq 0$ we have*

$$t\|\dot{x}(t)\|^2 \leq t\|T(x(t)) - x(t)\|^2 \leq \frac{2}{\underline{\tau}} \int_{t/2}^{t} \lambda(s)(1 - \lambda(s))\|T(x(s)) - x(s)\|^2 ds,$$

where $\underline{\tau} = \inf_{t \geq 0} \lambda(t)(1 - \lambda(t)) > 0$ and $\lim_{t \to +\infty} \int_{t/2}^{t} \lambda(s)(1 - \lambda(s))\|T(x(s)) - x(s)\|^2 ds = 0$.

Proof. Define the function $f : [0, +\infty) \to [0, +\infty)$,

$$f(t) = \int_0^t \lambda(s)(1 - \lambda(s))\|T(x(s)) - x(s)\|^2 ds.$$

According to (5. 4) we have that $\lim_{t \to +\infty} f(t) \in \mathbb{R}$.

Since $t \mapsto \frac{1}{2}\|T(x(t)) - x(t)\|^2$ is decreasing (see the proof of Theorem 5.1), we have for all $t \geq 0$:

$$\begin{aligned}\|T(x(t)) - x(t)\|^2 \int_{t/2}^{t} \lambda(s)(1 - \lambda(s))ds \leq & \int_{t/2}^{t} \lambda(s)(1 - \lambda(s))\|T(x(s)) - x(s)\|^2 ds \\ = & f(t) - f(t/2).\end{aligned}$$

Taking into account the definition of $\underline{\tau}$, we easily derive

$$\frac{\underline{\tau}}{2} t\|T(x(t)) - x(t)\|^2 \leq \int_{t/2}^{t} \lambda(s)(1 - \lambda(s))\|T(x(s)) - x(s)\|^2 ds,$$

and the conclusion follows by using again (5. 1). □

The remaining of the section is dedicated to the formulation and investigation of a continuous version of the forward-backward algorithm.

We need the following technical result regarding the averaged parameter of the composition of two averaged operators. We refer also to [26, Proposition 4.32] for other results of this type.

Proposition 5.1 *(see [111, Theorem 3(b)] and [77, Proposition 2.4]) Let $T_i : \mathcal{H} \to \mathcal{H}$ be α_i-averaged, where $\alpha_i \in (0, 1)$, $i = 1, 2$. Then the composition $T_1 \circ T_2$ is α-averaged, where*

$$\alpha = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \in (0, 1).$$

Theorem 5.4 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $\beta > 0$ and $B : \mathcal{H} \to \mathcal{H}$ be β-cocoercive such that* $\operatorname{zer}(A + B) \neq \emptyset$*. Let $\eta \in (0, 2\beta)$ and set $\delta = (4\beta - \eta)/(2\beta)$. Let $\lambda : [0, +\infty) \to [0, \delta]$ be a Lebesgue measurable function and $x_0 \in \mathcal{H}$. Suppose that one of the following conditions is fulfilled:*

$$\int_0^{+\infty} \lambda(t)(\delta - \lambda(t))dt = +\infty \ \textit{or} \ \inf_{t \geq 0} \lambda(t) > 0.$$

Let $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of

$$\begin{cases} \dot{x}(t) = \lambda(t) \left[J_{\eta A}\Big(x(t) - \eta B(x(t))\Big) - x(t) \right] \\ x(0) = x_0. \end{cases} \quad (5.\ 8)$$

Then the following statements are true:

(i) *the trajectory* x *is bounded and* $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$;

(ii) $\lim_{t\to+\infty} \Big[J_{\eta A}\Big(x(t) - \eta B(x(t))\Big) - x(t)\Big] = 0$;

(iii) $\lim_{t\to+\infty} \dot{x}(t) = 0$;

(iv) $x(t)$ *converges weakly to a point in* $\operatorname{zer}(A+B)$, *as* $t \to +\infty$.

Suppose that $\inf_{t\geq 0} \lambda(t) > 0$. *Then the following hold:*

(v) *if* $y \in \operatorname{zer}(A+B)$, *then* $\lim_{t\to+\infty} B(x(t)) = By$ *and* B *is constant on* $\operatorname{zer}(A+B)$;

(vi) *if* A *or* B *is uniformly monotone, then* $x(t)$ *converges strongly to the unique point in* $\operatorname{zer}(A+B)$, *as* $t \to +\infty$.

Proof. It is immediate that the dynamical system (5. 8) can be written in the form

$$\left\{ \begin{array}{l} \dot{x}(t) = \lambda(t)\big(T(x(t)) - x(t)\big) \\ x(0) = x_0, \end{array} \right. \tag{5. 9}$$

where $T = J_{\eta A} \circ (\operatorname{Id} - \eta B)$. According to [26, Corollary 23.8 and Remark 4.24(iii)], $J_{\eta A}$ is 1/2-cocoercive. Moreover, by [26, Proposition 4.33], $\operatorname{Id} - \eta B$ is $\eta/(2\beta)$-averaged. Combining this with Proposition 5.1, we derive that T is $1/\delta$-averaged. The statements (i)-(iv) follow now from Corollary 5.1 by noticing that $\operatorname{Fix} T = \operatorname{zer}(A+B)$, see [26, Proposition 25.1(iv)].

We suppose in the following that $\inf_{t\geq 0} \lambda(t) > 0$.

(v) The fact that B is constant on $\operatorname{zer}(A+B)$ follows from the cocoercivity of B and the monotonicity of A. A proof of this statement when A is the subdifferential of a proper, convex and lower semicontinuous function is given in [1, Lema 2.7].

We use the following inequality:

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 - \eta(2\beta - \eta)\|Bx - By\|^2 \ \forall (x, y) \in \mathcal{H} \times \mathcal{H}, \tag{5. 10}$$

which follows from the nonexpansiveness property of the resolvent and the cocoercivity of B:

$$\begin{aligned} \|Tx - Ty\|^2 &\leq \|x - y - \eta(Bx - By)\|^2 \\ &= \|x - y\|^2 - 2\eta \langle x - y, Bx - By\rangle + \eta^2\|Bx - By\|^2 \\ &\leq \|x - y\|^2 - \eta(2\beta - \eta)\|Bx - By\|^2. \end{aligned}$$

Take an arbitrary $x^* \in \operatorname{zer}(A+B) = \operatorname{Fix} T$. From the first part of the proof of Theorem 5.1 and (5. 10) we get for almost every $t \geq 0$

$$\begin{aligned} &\frac{d}{dt}\|x(t) - x^*\|^2 + \lambda(t)(1 - \lambda(t))\|T(x(t) - x(t))\|^2 + \|\dot{x}(t)\|^2 \\ &= \lambda(t)\|T(x(t)) - x^*\|^2 - \lambda(t)\|x(t) - x^*\|^2 \\ &\leq -\eta(2\beta - \eta)\lambda(t)\|B(x(t)) - Bx^*\|^2. \end{aligned}$$

Taking into account that $\inf_{t\geq 0} \lambda(t) > 0$ and $0 < \eta < 2\beta$, by integrating the above inequality we obtain

$$\int_0^{+\infty} \|B(x(t)) - Bx^*\|^2 dt < +\infty.$$

Since B is $1/\beta$-Lipschitz continuous (this follows from the β-cocoercivity of B by applying the Cauchy-Schwarz inequality) and $\|\dot{x}(\cdot)\| \in L^2([0, +\infty))$, from Remark

5.1(b) we derive that $t \mapsto \frac{d}{dt}B(x(t)) \in L^2([0,\infty),\mathcal{H})$. From the Cauchy-Schwarz inequality we obtain for all $t \geq 0$

$$\frac{d}{dt}\left(\|B(x(t)) - Bx^*\|^2\right) = 2\left\langle \frac{d}{dt}B(x(t)), B(x(t)) - Bx^*\right\rangle \leq \left\|\frac{d}{dt}B(x(t))\right\|^2 + \|B(x(t)) - Bx^*\|^2.$$

Combining these considerations with Lemma 5.2, we conclude that $B(x(t))$ converges strongly to Bx^*, as $t \to +\infty$.

(vi) Suppose that A is uniformly monotone and let x^* be the unique point in $\operatorname{zer}(A+B)$. According to (5. 8) and the definition of the resolvent, we have for almost every $t \geq 0$

$$-B(x(t)) - \frac{1}{\eta\lambda(t)}\dot{x}(t) \in A\left(\frac{1}{\lambda(t)}\dot{x}(t) + x(t)\right). \tag{5. 11}$$

From $-Bx^* \in Ax^*$ we get for almost every $t \geq 0$ the inequality

$$\phi_A\left(\left\|\frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|\right) \leq \left\langle \frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*, -B(x(t)) - \frac{1}{\eta\lambda(t)}\dot{x}(t) + Bx^*\right\rangle,$$

where $\phi_A : [0,+\infty) \to [0,+\infty]$ is increasing and vanishes only at 0.

The monotonicity of B implies

$$\begin{aligned}
&\phi_A\left(\left\|\frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|\right)\\
&\leq -\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{1}{\lambda(t)}\left\langle \dot{x}(t), -B(x(t)) + Bx^*\right\rangle\\
&\quad + \left\langle x(t) - x^*, -B(x(t)) + Bx^*\right\rangle - \frac{1}{\eta\lambda(t)}\left\langle \dot{x}(t), x(t) - x^*\right\rangle\\
&\leq -\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{1}{\lambda(t)}\left\langle \dot{x}(t), -B(x(t)) + Bx^*\right\rangle - \frac{1}{\eta\lambda(t)}\left\langle \dot{x}(t), x(t) - x^*\right\rangle.
\end{aligned}$$

The last inequality implies, by taking into consideration (iii), (iv) and (v), that

$$\lim_{t\to+\infty}\phi_A\left(\left\|\frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|\right) = 0.$$

The properties of the function ϕ_A allow to conclude that $\frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*$ converges strongly to 0, as $t \to +\infty$, hence from (iii) we obtain the conclusion.

Finally, suppose that B is uniformly monotone, with corresponding function $\phi_B : [0,+\infty) \to [0,+\infty]$, which is increasing and vanishes only at 0. The conclusion follows by taking in the inequality

$$\langle x(t) - x^*, B(x(t)) - Bx^*\rangle \geq \phi_B(\|x(t) - x^*\|)$$

the limit as $t \to +\infty$ and by using (i) and (v). □

Remark 5.6 We would like to emphasize the fact that the statements in Theorem 5.4 remain valid also for $\eta := 2\beta$. Indeed, in this case the cocoercivity of B implies that $\operatorname{Id} - \eta B$ is nonexpansive, hence the operator $T = J_{\eta A} \circ (\operatorname{Id} - \eta B)$ used in the proof is nonexpansive, too, and so the statements in (i)-(iv) follow from Theorem

5.1. Furthermore, for the proof of the statements (v) and (vi), the key observation was that $B(x(\cdot)) - Bx^* \in L^2([0,\infty),\mathcal{H})$, where $x^* \in \operatorname{zer}(A+B)$. Let us prove that this is true also in this case. Indeed, from (5. 11), the relation $-Bx^* \in Ax^*$ and the monotonicity of A we have for almost every $t \geq 0$ the inequality

$$0 \leq \left\langle \frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*, -B(x(t)) - \frac{1}{\eta\lambda(t)}\dot{x}(t) + Bx^* \right\rangle.$$

The cocoercivity of B implies

$$\begin{aligned}
0 \leq & -\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{1}{\lambda(t)}\left\langle \dot{x}(t), -B(x(t)) + Bx^* \right\rangle \\
& + \left\langle x(t) - x^*, -B(x(t)) + Bx^* \right\rangle - \frac{1}{\eta\lambda(t)}\left\langle \dot{x}(t), x(t) - x^* \right\rangle \\
\leq & \frac{1}{\lambda(t)}\left\langle \dot{x}(t), -B(x(t)) + Bx^* \right\rangle - \beta\|B(x(t)) - Bx^*\|^2 - \frac{1}{\eta\lambda(t)}\left\langle \dot{x}(t), x(t) - x^* \right\rangle \\
\leq & \frac{1}{2\beta\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{\beta}{2}\|B(x(t)) - Bx^*\|^2 \\
& - \beta\|B(x(t)) - Bx^*\|^2 - \frac{1}{\eta\lambda(t)}\frac{d}{dt}\left[\frac{1}{2}\|x(t) - x^*\|^2\right].
\end{aligned}$$

We derive that for almost every $t \geq 0$ the following inequality holds:

$$\frac{\beta\lambda(t)}{2}\|B(x(t)) - Bx^*\|^2 + \frac{1}{\eta}\frac{d}{dt}\left[\frac{1}{2}\|x(t) - x^*\|^2\right] \leq \frac{1}{2\beta\lambda(t)}\|\dot{x}(t)\|^2,$$

which in combination with (i), the assumption $\inf_{t\geq 0}\lambda(t) > 0$ and λ bounded above delivers the desired conclusion.

Remark 5.7 Let us mention that in case $A = \partial\Phi$, where $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function defined on a real Hilbert space $\mathcal{H}$, and for $\lambda(t) = 1$ for all $t \geq 0$, the dynamical system (5. 8) becomes

$$\left\{ \begin{array}{l} \dot{x}(t) + x(t) = \operatorname{prox}_{\eta\Phi}\left(x(t) - \eta B(x(t))\right) \\ x(0) = x_0, \end{array} \right. \tag{5. 12}$$

which has been studied in [1]. Notice that the weak convergence of the trajectories of (5. 12) is obtained in [1, Theorem 5.2] for a constant step-size $\eta \in (0, 4\beta)$.

Remark 5.8 The explicit discretization of (5. 8) with respect to the time variable t, with step size $h_n > 0$ and initial point x_0, yields the following iterative scheme:

$$\frac{x_{n+1} - x_n}{h_n} = \lambda_n\left[J_{\eta A}\Big(x_n - \eta Bx_n\Big) - x_n\right] \ \forall n \geq 0.$$

For $h_n = 1$ this becomes

$$x_{n+1} = x_n + \lambda_n\left[J_{\eta A}\Big(x_n - \eta Bx_n\Big) - x_n\right] \ \forall n \geq 0, \tag{5. 13}$$

which is the classical forward-backward algorithm for finding the set of zeros of $A+B$. Let us mention that the convergence of (5. 13) is guaranteed in [26, Theorem 25.8] under the condition

$$\sum_{n\in\mathbb{N}} \lambda_n(\delta' - \lambda_n) = +\infty,$$

where

$$\delta' = \min\left\{1, \frac{\beta}{\eta}\right\} + \frac{1}{2}. \tag{5. 14}$$

This is due to the fact that in the proof of [26, Theorem 25.8] one applies [26, Proposition 4.32] in order to show that $J_{\eta A}\circ(\mathrm{Id}-\eta B)$ is $1/\delta'$-averaged. However, as done in the proof above, one can apply [111, Theorem 3(b)] (see also [77, Proposition 2.4]) in order get a better parameter for the averaged operator $J_{\eta A}\circ(\mathrm{Id}-\eta B)$, namely $1/\delta=(2\beta)/(4\beta-\eta)$. Notice that under the hypothesis $0<\eta\leq 2\beta$ one can prove the following relation between the parameters mentioned above:

$$\delta'=\min\left\{1,\frac{\beta}{\eta}\right\}+\frac{1}{2}\leq(4\beta-\eta)/(2\beta)=\delta. \tag{5. 15}$$

Remark 5.9 As seen also in Section 3.2, the Douglas-Rachford algorithm for finding the set of zeros of the sum of two maximally monotone operators follows from the discrete version of the Krasnosel'skiĭ–Mann numerical scheme, see also [26]. Following the approach presented above, one can formulate a dynamical system of Douglas-Rachford-type, the existence and weak convergence of the trajectories being a consequence of the main results presented here. The same can be done for other iterative schemes which have their origins in the discrete Krasnosel'skiĭ–Mann algorithm, like are the generalized forward-backward splitting algorithm in [119] and the forward-Douglas-Rachford splitting algorithm in [61].

Time rescaling arguments

The aim of this subsection is to show that, by using time rescaling arguments as in [14], some of the asymptotic properties of the dynamical system (5. 1) can be derived from the one of an autonomous dynamical system governed by a cocoercive operator. Let us recall the following classical result, which can be deduced for example from [1, Theorem 4.1] by taking $\Phi=0$ as well as from Theorem 5.4 by choosing $Ax=0$ for all $x\in\mathcal{H}$ and $\lambda(t)=1$ for all $t\geq 0$.

Theorem 5.5 *Let $B:\mathcal{H}\to\mathcal{H}$ be a cocoercive operator such that $\operatorname{zer}B\neq\emptyset$ and $w_0\in\mathcal{H}$. Let $w:[0,+\infty)\to\mathcal{H}$ be the unique strong global solution of the dynamical system*

$$\left\{\begin{array}{l}\dot{w}(t)+B(w(t))=0\\ w(0)=w_0.\end{array}\right. \tag{5. 16}$$

Then the following statements are true:

(a) the trajectory w is bounded and $\int_0^{+\infty}\|\dot{w}(t)\|^2dt<+\infty$;

(b) $w(t)$ converges weakly to a point in $\operatorname{zer}B$, as $t\to+\infty$;

(c) $B(w(\cdot))$ converges strongly to 0, as $t\to+\infty$.

Let us consider again the dynamical system (5. 1), written in the form

$$\left\{\begin{array}{l}\dot{x}(t)+\lambda(t)(\mathrm{Id}-T)(x(t))=0\\ x(0)=x_0.\end{array}\right. \tag{5. 17}$$

We recall that T is nonexpansive such that $\operatorname{Fix}T\neq\emptyset$ and $\lambda:[0,\infty)\to[0,1]$ is Lebesgue measurable. By using a time rescaling argument as in [14, Lemma 4.1], we can prove a connection between the dynamical system (5. 17) and the system

$$\left\{\begin{array}{l}\dot{w}(t)+(\mathrm{Id}-T)(w(t))=0\\ w(0)=x_0.\end{array}\right. \tag{5. 18}$$

In the following we suppose that

$$\int_0^{+\infty}\lambda(t)dt=+\infty. \tag{5. 19}$$

Notice that the considerations which we make in the following remain valid also when one requires for the function λ an arbitrary positive upper bound. However, we choose as upper bound 1 in order to remain in the setting presented in the previous section.

Suppose that we have a solution w of (5. 18). By defining the function $T_1 : [0,\infty) \to [0,\infty)$, $T_1(t) = \int_0^t \lambda(s)ds$, one can easily see that $w \circ T_1$ is a solution of (5. 17).

Conversely, if x is a solution of (5. 17), then $x \circ T_2$ is a solution of (5. 18), where $T_2 : [0,+\infty) \to [0,+\infty)$ is defined implicitly as $\int_0^{T_2(t)} \lambda(s)ds = t$ (this is possible due to the properties of the the function λ).

In the arguments we used that

$$T_1'(t) = \lambda(t) \ \forall t \geq 0 \tag{5. 20}$$

and

$$T_2'(t)\lambda(T_2(t)) = 1 \ \forall t \geq 0. \tag{5. 21}$$

Further, since $B := \operatorname{Id} - T$ is 1/2-cocoercive (this follows from the nonexpansiveness of T), for the dynamical system (5. 18) one can apply the convergence results presented in Theorem 5.5. We would also like to notice that the existence of a strong global solution of (5. 1) follows from the corresponding result for (5. 18), while for the uniqueness property we have to make use of the considerations at the beginning of Section 5.1.1.

In the following we deduce the convergence statements of Theorem 5.1 from the one of Theorem 5.5 by using the time rescaling arguments presented above.

Let x be the unique strong global solution of (5. 1). Due to the uniqueness of the solutions of (5. 1) and (5. 18), we have $x = w \circ T_1$, where w is the unique strong global solution of (5. 18).

(i) From Theorem 5.5(a) we know that w is bounded, hence x is bounded, too. We have

$$\begin{aligned}\int_0^{+\infty} \|\dot{x}(s)\|^2 ds &= \lim_{t\to\infty} \int_0^t \|w'(T_1(s))\|^2 (\lambda(s))^2 ds \\ &\leq \lim_{t\to\infty} \int_0^t \|w'(T_1(s))\|^2 \lambda(s) ds \\ &= \lim_{t\to\infty} \int_0^{T_1(t)} \|w'(u)\|^2 du \\ &< +\infty,\end{aligned}$$

where we used Theorem 5.5(a) and the change of variables $T_1(s) = u$.

(ii) The statement follows from Theorem 5.5(c).

(iii) Is a direct consequence of the boundedness of λ, (ii) and the way the dynamic is defined.

(iv) From Theorem 5.5(b) it follows that $x(t) = w(T_1(t))$ converges weakly to a point in $\operatorname{zer} B = \operatorname{Fix} T$ as $t \to +\infty$.

Remark 5.10 In the light of the above considerations it follows that the conclusion of Theorem 5.1 remains valid also when assuming that $\int_0^{+\infty} \lambda(t)dt = +\infty$, which is a weaker condition than asking that $\int_0^{+\infty} \lambda(t)(1-\lambda(t))dt = +\infty$ or $\inf_{t\geq 0} \lambda(t) > 0$. A similar statement applies to Theorem 5.4, too. Notice also that the assumption that λ takes values in $[0,1]$, being strictly bounded away from the endpoints of this interval, was essential, in combination to the considerations made in the proof of Theorem 5.1, for deriving convergence rates for the trajectories of (5. 1). Finally, let us mention that, as pointed out in Remark 5.5, the assumption $\int_0^{+\infty} \lambda(t)(1-\lambda(t))dt = +\infty$ has a natural counterpart in the discrete case which guarantees convergence for the sequence of generated iterates, while this is not the case for the other two conditions on λ considered above.

5.1.2 Converges rates for strongly monotone inclusions

In this subsection we investigate the convergence rates of the trajectories of the continuous dynamical systems considered above in the strongly monotone case and strongly convex case, respectively, the later concerning convex optimization problems. In both cases, we obtain exponential convergence rates for the orbits.

The following result can bee seen as the continuous counterpart of [26, Proposition 25.9], where it is shown that the sequence iteratively generated by the forward-backward algorithm linearly converges to the unique solution of

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in Ax^* + Bx^*, \tag{5. 22}$$

provided that one of the two involved operators is strongly monotone.

Theorem 5.6 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $B : \mathcal{H} \to \mathcal{H}$ a monotone and $\frac{1}{\beta}$-Lipschitz continuous operator for $\beta > 0$ such that $A + B$ is ρ-strongly monotone for $\rho > 0$ and x^* be the unique point in $\operatorname{zer}(A + B)$. Let $\lambda : [0, +\infty) \to [0, +\infty)$ be a Lebesgue measurable function such that there exist real numbers $\underline{\lambda}$ and $\overline{\lambda}$ fulfilling*

$$0 < \underline{\lambda} \le \inf_{t\ge0} \lambda(t) \le \sup_{t\ge0} \lambda(t) \le \overline{\lambda}.$$

Chose $\alpha > 0$ and $\eta > 0$ such that

$$\alpha < 2\rho\beta^2\underline{\lambda} \text{ and } \frac{1}{\beta} + \frac{\overline{\lambda}}{2\alpha} \le \rho + \frac{1}{\eta}.$$

If $x_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ is the unique strong global solution of the dynamical system (5. 8), then for every $t \in [0, +\infty)$ one has

$$\|x(t) - x^*\|^2 \le \|x_0 - x^*\|^2 \exp(-Ct),$$

where $C := \frac{2\rho\underline{\lambda} - \frac{\alpha}{\beta^2}}{2\rho + \frac{1}{\eta}} > 0$.

Proof. Notice that B is a maximally monotone operator (see [26, Corollary 20.25]) and, since B has full domain, $A+B$ is maximally monotone, too (see [26, Corollary 24.4]). Therefore, due to the strong monotonicity of $A+B$, $\operatorname{zer}(A+B)$ is a singleton (see [26, Corollary 23.37]).

A direct consequence of (5. 8) and of the definition of the resolvent is the inclusion

$$-\frac{1}{\eta\lambda(t)}\dot{x}(t) - B(x(t)) + B\left(\frac{1}{\lambda(t)}\dot{x}(t) + x(t)\right) \in (A + B)\left(\frac{1}{\lambda(t)}\dot{x}(t) + x(t)\right),$$

which holds for almost every $t \in [0, +\infty)$. Combining it with $0 \in (A + B)(x^*)$ and the strong monotonicity of $A + B$, it yields for almost every $t \in [0, +\infty)$

$$\rho\left\|\frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|^2 \le$$
$$\left\langle \frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*, -\frac{1}{\eta\lambda(t)}\dot{x}(t) - B(x(t)) + B\left(\frac{1}{\lambda(t)}\dot{x}(t) + x(t)\right)\right\rangle.$$

By using the notation $h(t) = \frac{1}{2}\|x(t)-x^*\|^2$ for $t \in [0,+\infty)$, the Cauchy-Schwartz inequality, the Lipschitz property of B and the fact that $\dot{h}(t) = \langle x(t)-x^*, \dot{x}(t)\rangle$, we deduce that for almost every $t \in [0,+\infty)$

$$\begin{aligned}
&\rho\left\|\frac{1}{\lambda(t)}\dot{x}(t)+x(t)-x^*\right\|^2\\
\leq &-\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2+\frac{1}{\lambda(t)}\left\langle\dot{x}(t), B\left(\frac{1}{\lambda(t)}\dot{x}(t)+x(t)\right)-B(x(t))\right\rangle\\
&-\frac{1}{\eta\lambda(t)}\dot{h}(t)+\left\langle x(t)-x^*, B\left(\frac{1}{\lambda(t)}\dot{x}(t)+x(t)\right)-B(x(t))\right\rangle\\
\leq &-\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2+\frac{1}{\beta\lambda^2(t)}\|\dot{x}(t)\|^2-\frac{1}{\eta\lambda(t)}\dot{h}(t)\\
&+\frac{1}{\beta\lambda(t)}\|x(t)-x^*\|\|\dot{x}(t)\|\\
\leq &-\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2+\frac{1}{\beta\lambda^2(t)}\|\dot{x}(t)\|^2-\frac{1}{\eta\lambda(t)}\dot{h}(t)\\
&+\frac{\alpha}{\beta^2\lambda(t)}h(t)+\frac{1}{2\alpha\lambda(t)}\|\dot{x}(t)\|^2.
\end{aligned}$$

As

$$\rho\left\|\frac{1}{\lambda(t)}\dot{x}(t)+x(t)-x^*\right\|^2=\frac{\rho}{\lambda^2(t)}\|\dot{x}(t)\|^2+\frac{2\rho}{\lambda(t)}\dot{h}(t)+2\rho h(t),$$

we obtain for almost every $t \in [0,+\infty)$ the inequality

$$\begin{aligned}
&\left(\frac{2\rho}{\lambda(t)}+\frac{1}{\eta\lambda(t)}\right)\dot{h}(t)+\left(2\rho-\frac{\alpha}{\beta^2\lambda(t)}\right)h(t)+\\
&\left(\frac{\rho}{\lambda^2(t)}+\frac{1}{\eta\lambda^2(t)}-\frac{1}{\beta\lambda^2(t)}-\frac{1}{2\alpha\lambda(t)}\right)\|\dot{x}(t)\|^2\leq 0.
\end{aligned}$$

However, the way in which the involved parameters were chosen imply for almost every $t \in [0,+\infty)$ that

$$\left(\frac{2\rho}{\lambda(t)}+\frac{1}{\eta\lambda(t)}\right)\dot{h}(t)+\left(2\rho-\frac{\alpha}{\beta^2\lambda(t)}\right)h(t)\leq 0 \tag{5. 23}$$

or, equivalently,

$$\dot{h}(t)+\frac{2\rho\lambda(t)-\frac{\alpha}{\beta^2}}{2\rho+\frac{1}{\eta}}h(t)\leq 0.$$

This further implies

$$\dot{h}(t)+Ch(t)\leq 0$$

for almost every $t \in [0,+\infty)$. By multiplying this inequality with $\exp(Ct)$ and integrating from 0 to T, where $T \geq 0$, one easily obtains the conclusion. □

Remark 5.11

(a) By time rescaling arguments one could consider $\lambda(t) = 1$ for every $t \geq 0$ and, consequently, investigate the asymptotic properties of the system

$$\left\{\begin{array}{l}\dot{x}(t)+M(x(t))=0\\ x(0)=x_0,\end{array}\right. \tag{5. 24}$$

where $M : \mathcal{H} \to \mathcal{H}$ is defined by $M = \operatorname{Id} - J_{\eta A} \circ (\operatorname{Id} - \eta B)$. In the hypotheses of Theorem 5.6 the operator M satisfies the following inequality for all $x \in \mathcal{H}$:

$$\left(2\rho + \frac{1}{\eta}\right) \langle Mx, x - x^* \rangle \geq \left(\rho - \frac{\alpha}{2\beta^2}\right) \|x - x^*\|^2 + \left(\rho + \frac{1}{\eta} - \frac{1}{\beta} - \frac{1}{2\alpha}\right) \|Mx\|^2. \tag{5. 25}$$

This follows by using the same arguments as used in the proof of Theorem 5.6, namely the definition of the resolvent operator, the inclusion $0 \in (A+B)(x^*)$ and the strong monotonicity of $A + B$. Coming back to the system (5. 24), the exponential convergence rate for the trajectory is further obtained by applying the Gronwall Lemma in the inequality

$$\left(2\rho + \frac{1}{\eta}\right) \langle \dot{x}(t), x(t) - x^* \rangle + \left(\rho - \frac{\alpha}{2\beta^2}\right) \|x(t) - x^*\|^2 \leq 0,$$

which is nothing else than relation (5. 23) in the proof of Theorem 5.6.

(b) Notice that by chosing the involved parameters as in Theorem 5.6, relation (5. 25) yields the inequality

$$\left(2\rho + \frac{1}{\eta}\right) \langle Mx, x - x^* \rangle \geq \left(\rho - \frac{\alpha}{2\beta^2}\right) \|x - x^*\|^2 \ \forall x \in \mathcal{H},$$

where $Mx^* = 0$. Thus the operator M satisfies a strong monotone property in the sense of Pazy (see relation (11.2) in Theorem 11.2 in [114]). However, the hypotheses of Theorem 5.6 do not imply in general the strong monotonicity of the operator M in the sense of (1. 31), thus the result presented in Theorem 5.6 does not fall into the framework of the classical result concerning exponential convergence rates for the semigroup generated by a strongly monotone operator as presented in [60, Theorem 3.9].

Further, we discuss some situations when the operator M is strongly monotone in the classical sense (see (1. 31)). We start with two trivial cases. The first one is $Ax = 0$ for all $x \in \mathcal{H}$ and B is strongly monotone. The second one is $Bx = 0$ for all $x \in \mathcal{H}$ and A is strongly monotone, in which case $J_{\eta A}$ is a contraction (see [26, Proposition 23.11]), hence $M = \operatorname{Id} - J_{\eta A}$ is strongly monotone. Other situations follow in the framework of [26, Proposition 25.9]: i) if A is strongly monotone, B is β-cocoercive and $\eta < 2\beta$; ii) if B is θ-strongly monotone and β^{-1}-Lipschitz continuous, $\theta\beta \leq 1$ and $\eta < 2\theta\beta^2$.

We come now to the convex optimization problem

$$\min_{x \in \mathcal{H}} \ f(x) + g(x), \tag{5. 26}$$

where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ is a convex and (Fréchet) differentiable function with Lipschitz continuous gradient. Notice that, since

$$\operatorname{argmin}(f + g) = \operatorname{zer}(\partial(f + g)) = \operatorname{zer}(\partial f + \nabla g),$$

one can approach this set by means of the trajectories of the dynamical system (5. 8) written for $A = \partial f$ and $B = \nabla g$. This being said, the dynamical system (5. 8) becomes

$$\left\{ \begin{array}{l} \dot{x}(t) = \lambda(t) \left[\operatorname{prox}_{\eta f} \Big(x(t) - \eta \nabla g(x(t))\Big) - x(t)\right] \\ x(0) = x_0. \end{array} \right. \tag{5. 27}$$

The following result is a direct consequence of Theorem 5.6. Let us also notice that in case $f + g$ is ρ-strongly convex for $\rho > 0$, the operator $\partial(f + g) = \partial f + \nabla g$ is a ρ-strongly monotone operator (see [26, Example 22.3(iv)].)

Theorem 5.7 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function, $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\beta > 0$ such that $f + g$ is ρ-strongly convex for $\rho > 0$ and x^* be the unique minimizer of $f+g$ over $\mathcal{H}$. Let $\lambda : [0, +\infty) \to [0, +\infty)$ be a Lebesgue measurable function such that there exist real numbers $\underline{\lambda}$ and $\overline{\lambda}$ fulfilling*

$$0 < \underline{\lambda} \leq \inf_{t \geq 0} \lambda(t) \leq \sup_{t \geq 0} \lambda(t) \leq \overline{\lambda}.$$

Chose $\alpha > 0$ and $\eta > 0$ such that

$$\alpha < 2\rho\beta^2\underline{\lambda} \text{ and } \frac{1}{\beta} + \frac{\overline{\lambda}}{2\alpha} \leq \rho + \frac{1}{\eta}.$$

If $x_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ is the unique strong global solution of the dynamical system (5. 27), then for every $t \in [0. + \infty)$ one has

$$\|x(t) - x^*\|^2 \leq \|x_0 - x^*\|^2 \exp(-Ct),$$

where $C := \frac{2\rho\underline{\lambda} - \frac{\alpha}{\beta^2}}{2\rho + \frac{1}{\eta}} > 0$.

In the last part of this section we approach the convex minimization problem

$$\min_{x \in \mathcal{H}} g(x), \tag{5. 28}$$

via the first order dynamical system

$$\left\{ \begin{array}{l} \dot{x}(t) + \lambda(t)\nabla g(x(t)) = 0 \\ x(0) = x_0. \end{array} \right. \tag{5. 29}$$

The following result quantifies the rate of convergence of g to its minimum value along the trajectories generated by (5. 29).

Theorem 5.8 *Let $g : \mathcal{H} \to \mathbb{R}$ be a ρ-strongly convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\rho > 0$ and $\beta > 0$ and x^* be the unique minimizer of g over $\mathcal{H}$. Let $\lambda : [0, +\infty) \to [0, +\infty)$ be a Lebesgue measurable function such that $\lambda(\cdot) \in L^1_{\text{loc}}[0, +\infty)$ and there exists a real number $\underline{\lambda} \in \mathbb{R}$ fulfilling*

$$0 < \underline{\lambda} \leq \inf_{t \geq 0} \lambda(t).$$

Chose $\alpha > 0$ such that

$$\alpha \leq 2\underline{\lambda}\beta\rho^2.$$

If $x_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ is the unique strong global solution of the dynamical system (5. 29), then for every $t \in [0, +\infty)$ one has

$$\begin{aligned} 0 &\leq \frac{\rho}{2}\|x(t) - x^*\|^2 \\ &\leq g(x(t)) - g(x^*) \\ &\leq (g(x_0) - g(x^*)) \exp(-\alpha t) \\ &\leq \frac{1}{2\beta}\|x_0 - x^*\|^2 \exp(-\alpha t). \end{aligned}$$

Proof. The second inequality is a consequence of the strong convexity of the function g. Further, by noticing that $\nabla g(x^*) = 0$, from Lemma 1.4 we obtain

$$g(x(t)) - g(x^*) \leq \frac{1}{2\beta}\|x(t) - x^*\|^2. \tag{5. 30}$$

From here, the last inequality in the conclusion follows automatically.

Using the strong convexity of g we have for every $t \in [0, +\infty)$ that

$$\rho\|x(t) - x^*\|^2 \leq \langle x(t) - x^*, \nabla g(x(t))\rangle \leq \|x(t) - x^*\|\|\nabla g(x(t))\|,$$

thus

$$\rho\|x(t) - x^*\| \leq \|\nabla g(x(t))\|. \tag{5. 31}$$

Finally, from the first equation in (5. 29), (5. 30), (5. 31) and using the way in which α was chosen, we obtain for almost every $t \in [0, +\infty)$

$$\begin{aligned}\frac{d}{dt}\big(g(x(t)) - g(x^*)\big) + \alpha\big(g(x(t)) - g(x^*)\big) &= \langle \dot{x}(t), \nabla g(x(t))\rangle + \alpha(g(x(t)) - g(x^*)) \\ &\leq -\lambda(t)\|\nabla g(x(t))\|^2 + \frac{\alpha}{2\beta}\|x(t) - x^*\|^2 \\ &\leq \left(-\lambda(t) + \frac{\alpha}{2\beta\rho^2}\right)\|\nabla g(x(t))\|^2 \\ &\leq 0.\end{aligned}$$

By multiplying this inequality with $\exp(\alpha t)$ and integrating from 0 to T, where $T \geq 0$, one easily obtains also the third inequality. $\square$

5.2 Second order dynamical systems

In this section we investigate the asymptotic behavior of the trajectories of second order dynamical systems associated to monotone inclusion problems.

5.2.1 Second order dynamical systems for monotone inclusion problems

Let us start with the study of existence and uniqueness of strong global solutions of a second order dynamical system governed by Lipschitz continuous operators.

Let $\Gamma : \mathcal{H} \to \mathcal{H}$ be an L_Γ-Lipschitz continuous operator, with $L_\Gamma \geq 0$, $B : \mathcal{H} \to \mathcal{H}$ be L_B-Lipschitz continuous, with $L_B \geq 0$, $\lambda : [0, +\infty) \to [0, +\infty)$ a Lebesgue measurable function, $u_0, v_0 \in \mathcal{H}$ and consider the dynamical system

$$\left\{\begin{array}{l}\ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)B(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0.\end{array}\right. \tag{5. 32}$$

Definition 5.3 We say that $x : [0, +\infty) \to \mathcal{H}$ is a strong global solution of (5. 32) if the following properties are satisfied:

(i) $x, \dot{x} : [0, +\infty) \to \mathcal{H}$ are locally absolutely continuous, in other words, absolutely continuous on each interval $[0, b]$ for $0 < b < +\infty$;

(ii) $\ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)B(x(t)) = 0$ for almost every $t \in [0, +\infty)$;

(iii) $x(0) = u_0$ and $\dot{x}(0) = v_0$.

For proving the existence and uniqueness of strong global solutions of (5. 32) we use the Cauchy-Lipschitz-Picard Theorem for absolutely continues trajectories (see for example [90, Proposition 6.2.1], [125, Theorem 54]). The key observation here is that one ca rewrite (5. 32) as a certain first order dynamical system in a product space (see also [6]).

Theorem 5.9 *Let $\Gamma : \mathcal{H} \to \mathcal{H}$ be an L_Γ-Lipschitz continuous operator, $B : \mathcal{H} \to \mathcal{H}$ a L_B-Lipschitz continuous operator and $\lambda : [0, +\infty) \to [0, +\infty)$ a Lebesgue measurable function such that $\lambda \in L^1_{\text{loc}}([0, +\infty))$ (that is $\lambda \in L^1([0, b])$ for every $0 < b < +\infty$). Then for each $u_0, v_0 \in \mathcal{H}$ there exists a unique strong global solution of the dynamical system* (5. 32).

Proof. The system (5. 32) can be equivalently written as a first order dynamical system in the phase space $\mathcal{H} \times \mathcal{H}$

$$\begin{cases} \dot{Y}(t) = F(t, Y(t)) \\ Y(0) = (u_0, v_0), \end{cases} \tag{5. 33}$$

with

$$Y : [0, +\infty) \to \mathcal{H} \times \mathcal{H},\ Y(t) = (x(t), \dot{x}(t))$$

and

$$F : [0, +\infty) \times \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H},\ F(t, u, v) = (v, -\Gamma v - \lambda(t)Bu).$$

We endow $\mathcal{H} \times \mathcal{H}$ with scalar product

$$\langle (u, v), (\overline{u}, \overline{v}) \rangle_{\mathcal{H} \times \mathcal{H}} = \langle u, \overline{u} \rangle + \langle v, \overline{v} \rangle$$

and corresponding norm

$$\|(u, v)\|_{\mathcal{H} \times \mathcal{H}} = \sqrt{\|u\|^2 + \|v\|^2}.$$

(a) For arbitrary $u, \overline{u}, v, \overline{v} \in \mathcal{H}$, by using the Lipschitz continuity of the involved operators, we obtain for all $t \geq 0$:

$$\begin{aligned} \|F(t, u, v) - F(t, \overline{u}, \overline{v})\|_{\mathcal{H} \times \mathcal{H}} &= \sqrt{\|v - \overline{v}\|^2 + \|\Gamma\overline{v} - \Gamma v + \lambda(t)(B\overline{u} - Bu)\|^2} \\ &\leq \sqrt{(1 + 2L_\Gamma^2)\|v - \overline{v}\|^2 + 2L_B^2\lambda^2(t)\|u - \overline{u}\|^2} \\ &\leq \sqrt{1 + 2L_\Gamma^2 + 2L_B^2\lambda^2(t)}\|(u, \overline{u}) - (v, \overline{v})\|_{\mathcal{H} \times \mathcal{H}} \\ &\leq (1 + \sqrt{2}L_\Gamma + \sqrt{2}L_B\lambda(t))\|(u, \overline{u}) - (v, \overline{v})\|_{\mathcal{H} \times \mathcal{H}}. \end{aligned}$$

As $\lambda \in L^1_{\text{loc}}([0, +\infty))$, the Lipschitz constant of $F(t, \cdot, \cdot)$ is local integrable.

(b) Next we show that

$$\forall u, v \in \mathcal{H},\ \forall b > 0, \quad F(\cdot, u, v) \in L^1([0, b], \mathcal{H} \times \mathcal{H}). \tag{5. 34}$$

For arbitrary $u, v \in \mathcal{H}$ and $b > 0$ it holds

$$\begin{aligned} \int_0^b \|F(t, u, v)\|_{\mathcal{H} \times \mathcal{H}} dt &= \int_0^b \sqrt{\|v\|^2 + \|\Gamma v + \lambda(t)Bu\|^2} dt \\ &\leq \int_0^b \sqrt{\|v\|^2 + 2\|\Gamma v\|^2 + 2\lambda^2(t)\|Bu\|^2} dt \\ &\leq \int_0^b \left(\sqrt{\|v\|^2 + 2\|\Gamma v\|^2} + \sqrt{2}\lambda(t)\|Bu\| \right) dt \end{aligned}$$

and from here (5. 34) follows, by using the assumptions made on λ.

In the light of the statements (a) and (b), the existence and uniqueness of a strong global solution for (5. 33) are consequences of the Cauchy-Lipschitz-Picard Theorem for first order dynamical systems (see, for example, [90, Proposition 6.2.1], [125, Theorem 54]). From here, due to the equivalence of (5. 32) and (5. 33), the conclusion follows. $\square$

In the following we address the convergence properties of the trajectories generated by the dynamical system (5. 32) by assuming that $B : \mathcal{H} \to \mathcal{H}$ is a β-*cocoercive* operator for $\beta > 0$.

In order to prove the convergence of the trajectories of (5. 32), we make the following assumptions on the operator Γ and the relaxation function λ, respectively:

(A1) $\Gamma : \mathcal{H} \to \mathcal{H}$ is a bounded self-adjoint linear operator, assumed to be elliptic, that is, there exists $\gamma > 0$ such that $\langle \Gamma u, u\rangle \geq \gamma \|u\|^2$ for all $u \in \mathcal{H}$;

(A2) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\lambda}(t) \geq 0 \text{ and } \lambda(t) \leq \frac{\beta\gamma^2}{1+\theta}. \tag{5. 35}$$

Due to Definition 5.1 and Remark 5.1(a) $\dot{\lambda}(t)$ exists for almost every $t \geq 0$ and $\dot{\lambda}$ is Lebesgue integrable on each interval $[0, b]$ for $0 < b < +\infty$. If $\dot{\lambda}(t) \geq 0$ for almost every $t \geq 0$, then λ is monotonically increasing, thus, as λ is assumed to take only positive values, (A2) yields the existence of a lower bound $\underline{\lambda}$ such that for almost every $t \in [0, +\infty)$ one has

$$0 < \underline{\lambda} \leq \lambda(t) \leq \frac{\beta\gamma^2}{1+\theta}. \tag{5. 36}$$

We would also like to point out that under the conditions considered in (A2) the global version of the Picard-Lindelöf Theorem allows us to conclude that, for $u_0, v_0 \in \mathcal{H}$, there exists a unique trajectory $x : [0, +\infty) \to \mathcal{H}$ which is a C^2 function and which satisfies the relation (ii) in Definition 5.3 for every $t \in [0, +\infty)$. The considerations we make in the following take into account this fact.

Theorem 5.10 *Let* $B : \mathcal{H} \to \mathcal{H}$ *be a* β*-cocoercive operator for* $\beta > 0$ *such that* $\operatorname{zer} B \neq \emptyset$, $\Gamma : \mathcal{H} \to \mathcal{H}$ *be an operator fulfilling* (A1), $\lambda : [0, +\infty) \to (0, +\infty)$ *be a function fulfilling* (A2) *and* $u_0, v_0 \in \mathcal{H}$. *Let* $x : [0, +\infty) \to \mathcal{H}$ *be the unique strong global solution of* (5. 32). *Then the following statements are true:*

(i) the trajectory x *is bounded and* $\dot{x}, \ddot{x}, Bx \in L^2([0, +\infty); \mathcal{H})$*;*

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} B(x(t)) = 0$*;*

(iii) $x(t)$ *converges weakly to an element in* $\operatorname{zer} B$ *as* $t \to +\infty$.

Proof. Notice that the existence and uniqueness of the trajectory x follows from Theorem 5.9, since B is $1/\beta$-Lipschitz continuous, Γ is $\|\Gamma\|$-Lipschitz continuous and (A2) ensures $\lambda(\cdot) \in L^1_{\text{loc}}([0, +\infty))$.

(i) Take an arbitrary $x^* \in \operatorname{zer} B$ and consider for every $t \in [0, +\infty)$ the function

$$h(t) = \frac{1}{2}\|x(t) - x^*\|^2.$$

We have

$$\dot{h}(t) = \langle x(t) - x^*, \dot{x}(t)\rangle$$

and

$$\ddot{h}(t) = \|\dot{x}(t)\|^2 + \langle x(t) - x^*, \ddot{x}(t)\rangle,$$

for all $t \in [0, +\infty)$. Taking into account (5. 32), we get for all $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma\dot{h}(t) + \lambda(t)\langle x(t) - x^*, B(x(t))\rangle + \langle x(t) - x^*, \Gamma(\dot{x}(t)) - \gamma\dot{x}(t)\rangle = \|\dot{x}(t)\|^2. \tag{5. 37}$$

Now we introduce the function $p : [0, +\infty) \to \mathbb{R}$,

$$p(t) = \frac{1}{2} \langle (\Gamma - \gamma \operatorname{Id})(x(t) - x^*), x(t) - x^* \rangle . \tag{5. 38}$$

Due to (A1), as $\langle (\Gamma - \gamma \operatorname{Id}) u, u \rangle \geq 0$ for all $u \in \mathcal{H}$, it holds

$$p(t) \geq 0 \text{ for all } t \geq 0. \tag{5. 39}$$

Moreover,

$$\dot{p}(t) = \langle (\Gamma - \gamma \operatorname{Id})(\dot{x}(t)), x(t) - x^* \rangle ,$$

which combined with (5. 37), the cocoercivity of B and the fact that $Bx^* = 0$ yields for all $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma \dot{h}(t) + \beta \lambda(t) \|B(x(t))\|^2 + \dot{p}(t) \leq \|\dot{x}(t)\|^2.$$

Taking into account (5. 32) one obtains for all $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma \dot{h}(t) + \frac{\beta}{\lambda(t)} \|\ddot{x}(t) + \Gamma(\dot{x}(t))\|^2 + \dot{p}(t) \leq \|\dot{x}(t)\|^2,$$

hence

$$\ddot{h}(t) + \gamma \dot{h}(t) + \frac{\beta}{\lambda(t)} \|\ddot{x}(t)\|^2 + \frac{2\beta}{\lambda(t)} \langle \ddot{x}(t), \Gamma(\dot{x}(t)) \rangle + \frac{\beta}{\lambda(t)} \|\Gamma(\dot{x}(t))\|^2 + \dot{p}(t) \leq \|\dot{x}(t)\|^2. \tag{5. 40}$$

According to (A1) we have

$$\gamma \|u\| \leq \|\Gamma u\| \text{ for all } u \in \mathcal{H}, \tag{5. 41}$$

which combined with (5. 40) yields for all $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma \dot{h}(t) + \dot{p}(t) + \frac{\beta}{\lambda(t)} \frac{d}{dt} \big(\langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \big) + \left(\frac{\beta \gamma^2}{\lambda(t)} - 1 \right) ||\dot{x}(t)||^2 + \frac{\beta}{\lambda(t)} ||\ddot{x}(t)||^2 \leq 0.$$

By taking into account that for almost every $t \in [0, +\infty)$

$$\begin{aligned} \frac{1}{\lambda(t)} \frac{d}{dt} \big(\langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \big) = \ & \frac{d}{dt} \left(\frac{1}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \right) + \frac{\dot{\lambda}(t)}{\lambda^2(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \\ \geq \ & \frac{d}{dt} \left(\frac{1}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \right) + \gamma \frac{\dot{\lambda}(t)}{\lambda^2(t)} \|\dot{x}(t)\|^2, \end{aligned} \tag{5. 42}$$

we obtain for all $t \in [0, +\infty)$

$$\begin{aligned} & \ddot{h}(t) + \gamma \dot{h}(t) + \dot{p}(t) + \\ & \beta \frac{d}{dt} \left(\frac{1}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \right) + \left(\frac{\beta \gamma^2}{\lambda(t)} + \beta \gamma \frac{\dot{\lambda}(t)}{\lambda^2(t)} - 1 \right) ||\dot{x}(t)||^2 + \frac{\beta}{\lambda(t)} ||\ddot{x}(t)||^2 \leq 0. \end{aligned} \tag{5. 43}$$

By using now assumption (A2) we obtain that the following inequality holds for almost every $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma \dot{h}(t) + \dot{p}(t) + \beta \frac{d}{dt} \left(\frac{1}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \right) + \theta ||\dot{x}(t)||^2 + \frac{1 + \theta}{\gamma^2} ||\ddot{x}(t)||^2 \leq 0. \tag{5. 44}$$

This implies that the function $t \mapsto \dot{h}(t) + \gamma h(t) + p(t) + \frac{\beta}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle$, which is locally absolutely continuous, is monotonically decreasing. Hence there exists a real number M such that for all $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma h(t) + p(t) + \frac{\beta}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \leq M, \tag{5. 45}$$

which yields, together with (5. 39) and (A2), that for all $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma h(t) \leq M.$$

By multiplying this inequality with $e^{\gamma t}$ and then integrating from 0 to T, where $T > 0$, one easily obtains

$$h(T) \leq h(0)e^{-\gamma T} + \frac{M}{\gamma}(1 - e^{-\gamma T}),$$

thus

$$h \text{ is bounded} \tag{5. 46}$$

and, consequently,

$$\text{the trajectory } x \text{ is bounded.} \tag{5. 47}$$

On the other hand, from (5. 45), by taking into account (5. 39), (A1) and (A2), it follows that for all $t \in [0, +\infty)$

$$\dot{h}(t) + \frac{1+\theta}{\gamma} \|\dot{x}(t)\|^2 \leq M,$$

hence

$$\langle x(t) - x^*, \dot{x}(t) \rangle + \frac{1+\theta}{\gamma} \|\dot{x}(t)\|^2 \leq M.$$

This inequality, in combination with (5. 47), yields

$$\dot{x} \text{ is bounded,} \tag{5. 48}$$

which further implies that

$$\dot{h} \text{ is bounded.} \tag{5. 49}$$

Integrating the inequality (5. 44) we obtain that there exists a real number $N \in \mathbb{R}$ such that for all $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma h(t) + p(t) + \frac{\beta}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle + \theta \int_0^t ||\dot{x}(s)||^2 ds + \frac{1+\theta}{\gamma^2} \int_0^t ||\ddot{x}(s)||^2 ds \leq N.$$

From here, via (5. 49), (5. 39) and (A1), we conclude that $\dot{x}(\cdot), \ddot{x}(\cdot) \in L^2([0, +\infty); \mathcal{H})$. Finally, from (5. 32), (A1) and (A2) we deduce $Bx \in L^2([0, +\infty); \mathcal{H})$ and the proof of (i) is complete.

(ii) For all $t \in [0, +\infty)$ it holds

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t)\|^2\right) = \langle \dot{x}(t), \ddot{x}(t) \rangle \leq \frac{1}{2}\|\dot{x}(t)\|^2 + \frac{1}{2}\|\ddot{x}(t)\|^2$$

and Lemma 5.2 together with (i) lead to $\lim_{t \to +\infty} \dot{x}(t) = 0$.

Further, by taking into consideration Remark 5.1(b), for all $t \in [0, +\infty)$ we have

$$\frac{d}{dt}\left(\frac{1}{2}\|B(x(t))\|^2\right) = \left\langle B(x(t)), \frac{d}{dt}(Bx(t)) \right\rangle \leq \frac{1}{2}\|B(x(t))\|^2 + \frac{1}{2\beta^2}\|\dot{x}(t)\|^2.$$

By using again Lemma 5.2 and (i) we get $\lim_{t\to+\infty} B(x(t)) = 0$, while the fact that $\lim_{t\to+\infty} \ddot{x}(t) = 0$ follows from (5. 32), (A1) and (A2).

(iii) As seen in the proof of part (i), the function $t \mapsto \dot{h}(t) + \gamma h(t) + p(t) + \frac{\beta}{\lambda(t)}\langle \dot{x}(t), \Gamma(\dot{x}(t))\rangle$ is monotonically decreasing, thus from (i), (ii), (5. 39), (A1) and (A2) we deduce that $\lim_{t\to+\infty}(\gamma h(t) + p(t))$ exists and it is a real number.

In the following we consider the scalar product defined by $\langle\langle x, y\rangle\rangle = \frac{1}{\gamma}\langle \Gamma x, y\rangle$ and the corresponding induced norm $|||x|||^2 = \frac{1}{\gamma}\langle \Gamma x, x\rangle$. Taking into account the definition of p, we have that $\lim_{t\to+\infty} \frac{1}{2}|||x(t) - x^*|||^2$ exists and it is a real number.

Let $\overline{x}$ be a weak sequential cluster point of x, that is, there exists a sequence $t_n \to +\infty$ (as $n \to +\infty$) such that $(x(t_n))_{n\in\mathbb{N}}$ converges weakly to $\overline{x}$. Since B is a maximally monotone operator (see for instance [26, Example 20.28]), its graph is sequentially closed with respect to the weak-strong topology of the product space $\mathcal{H}\times\mathcal{H}$. By using also that $\lim_{n\to+\infty} B(x(t_n)) = 0$, we conclude that $B\overline{x} = 0$, hence $\overline{x} \in \operatorname{zer} B$.

The conclusion follows by applying the Opial Lemma 1.1 in the Hilbert space $(\mathcal{H}, (\langle\langle\cdot,\cdot\rangle\rangle))$, by noticing that a sequence $(x_n)_{n\geq 0}$ converges weakly to $\overline{x} \in \mathcal{H}$ in $(\mathcal{H}, (\langle\langle\cdot,\cdot\rangle\rangle))$ if and only if $(x_n)_{n\geq 0}$ converges weakly to $\overline{x}$ in $(\mathcal{H}, (\langle\cdot,\cdot\rangle))$. $\square$

A standard instance of a cocoercive operator defined on a real Hilbert spaces is the one that can be represented as $B = \operatorname{Id} - T$, where $T : \mathcal{H} \to \mathcal{H}$ is a nonexpansive operator. As it easily follows from the nonexpansiveness of T, B is in this case 1/2-cocoercive. For this particular choice of the operator B, the dynamical system (5. 32) becomes

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\big(x(t) - T(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{5. 50}$$

while assumption (A2) reads

(A3) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\lambda}(t) \geq 0 \text{ and } \lambda(t) \leq \frac{\gamma^2}{2(1+\theta)}. \tag{5. 51}$$

Theorem 5.10 gives rise to the following result.

Corollary 5.2 *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator such that* $\operatorname{Fix} T = \{u \in \mathcal{H} : Tu = u\} \neq \emptyset$, *$\Gamma : \mathcal{H} \to \mathcal{H}$ be an operator fulfilling* (A1), *$\lambda : [0, +\infty) \to (0, +\infty)$ be a function fulfilling* (A3) *and $u_0, v_0 \in \mathcal{H}$. Let $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 50). *Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, (\operatorname{Id} - T)x \in L^2([0, +\infty); \mathcal{H})$*;*

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty}(\operatorname{Id} - T)(x(t)) = 0$*;*

(iii) $x(t)$ converges weakly to a point in $\operatorname{Fix} T$ *as $t \to +\infty$.*

Remark 5.12 In the particular case when $\Gamma = \gamma \operatorname{Id}$ for $\gamma > 0$ and $\lambda(t) = 1$ for all $t \in [0, +\infty)$ the dynamical system (5. 50) becomes

$$\begin{cases} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) - T(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{5. 52}$$

The convergence of the trajectories generated by (5. 52) has been studied in [8, Theorem 3.2] under the condition $\gamma^2 > 2$. In this case (A3) is obviously fulfilled for an arbitrary $0 < \theta \leq (\gamma^2 - 2)/2$. However, different to [8], we allow in Corollary 5.2 an anisotropic damping through the use of the elliptic operator Γ and also a variable relaxation function λ depending on time (in [3] the anisotropic damping has been considered as well in the context of minimizing of a smooth convex function via second order dynamical systems).

We close the section by addressing an immediate consequence of the above corollary applied to second order dynamical systems governed by averaged operators.

We consider the dynamical system

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\big(x(t) - R(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0 \end{cases} \tag{5. 53}$$

and formulate the assumption

(A4) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\lambda}(t) \geq 0 \text{ and } \lambda(t) \leq \frac{\gamma^2}{2\alpha(1+\theta)}. \tag{5. 54}$$

Corollary 5.3 *Let $R : \mathcal{H} \to \mathcal{H}$ be an α-averaged operator for $\alpha \in (0,1)$ such that $\operatorname{Fix} R \neq \emptyset$, $\Gamma : \mathcal{H} \to \mathcal{H}$ be an operator fulfilling* (A1)*, $\lambda : [0, +\infty) \to (0, +\infty)$ be a function fulfilling* (A4) *and $u_0, v_0 \in \mathcal{H}$. Let $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 53)*. Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, (\operatorname{Id} - R)x \in L^2([0, +\infty); \mathcal{H})$;

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} (\operatorname{Id} - R)(x(t)) = 0$*;*

(iii) $x(t)$ converges weakly to a point in $\operatorname{Fix} R$ as $t \to +\infty$.

Proof. Since R is α-averaged, there exists a nonexpansive operator $T : \mathcal{H} \to \mathcal{H}$ such that $R = (1-\alpha)\operatorname{Id} + \alpha T$. The conclusion is a direct consequence of Corollary 5.2, by taking into account that (5. 53) is equivalent to

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \alpha\lambda(t)\big(x(t) - T(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases}$$

and $\operatorname{Fix} R = \operatorname{Fix} T$. □

5.2.2 Second order dynamical systems of forward-backward type

In this section we address the monotone inclusion problem

$$\text{find } x^* \in \mathcal{H} \text{ such that } \ 0 \in A(x^*) + B(x^*),$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ is a β-cocoercive operator for $\beta > 0$ via a second order forward-backward dynamical system with anisotropic damping and variable relaxation parameter.

For $\eta > 0$ we consider the dynamical system

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\Big[x(t) - J_{\eta A}\Big(x(t) - \eta B(x(t))\Big)\Big] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{5. 55}$$

We formulate the following assumption, where $\delta := (4\beta - \eta)/(2\beta)$:

(A5) $\lambda : [0,+\infty) \to (0,+\infty)$ is locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0,+\infty)$ we have

$$\dot{\lambda}(t) \geq 0 \text{ and } \lambda(t) \leq \frac{\delta\gamma^2}{2(1+\theta)}. \tag{5. 56}$$

Theorem 5.11 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ be β-cocoercive operator for $\beta > 0$ such that $\operatorname{zer}(A+B) \neq \emptyset$. Let $\eta \in (0,2\beta)$ and set $\delta := (4\beta-\eta)/(2\beta)$. Let $\Gamma : \mathcal{H} \to \mathcal{H}$ be an operator fulfilling* (A1)*, $\lambda : [0,+\infty) \to (0,+\infty)$ be a function fulfilling* (A5)*, $u_0, v_0 \in \mathcal{H}$ and $x : [0,+\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 55)*. Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, \big(\operatorname{Id} - J_{\eta A} \circ (\operatorname{Id} - \eta B)\big)x \in L^2([0,+\infty);\mathcal{H})$;

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} \big(\operatorname{Id} - J_{\eta A} \circ (\operatorname{Id} - \eta B)\big)(x(t)) = 0$;

(iii) $x(t)$ converges weakly to a point in $\operatorname{zer}(A+B)$ as $t \to +\infty$;

(iv) if $x^* \in \operatorname{zer}(A+B)$, then $B(x(\cdot)) - Bx^* \in L^2([0,+\infty);\mathcal{H})$, $\lim_{t\to+\infty} B(x(t)) = Bx^*$ and B is constant on $\operatorname{zer}(A+B)$;

(v) if A or B is uniformly monotone, then $x(t)$ converges strongly to the unique point in $\operatorname{zer}(A+B)$ as $t \to +\infty$.

Proof. (i)-(iii) It is immediate that the dynamical system (5. 55) can be written in the form

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\big(x(t) - R(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{5. 57}$$

where $R = J_{\eta A} \circ (\operatorname{Id} - \eta B)$. According to [26, Corollary 23.8 and Remark 4.24(iii)], $J_{\eta A}$ is 1/2-cocoercive. Moreover, by [26, Proposition 4.33], $\operatorname{Id} - \eta B$ is $\eta/(2\beta)$-averaged. Combining this with Proposition 5.1, we derive that R is $1/\delta$-averaged. The statements (i)-(iii) follow now from Corollary 5.3 by noticing that $\operatorname{Fix} R = \operatorname{zer}(A+B)$ (see [26, Proposition 25.1(iv)]).

(iv) The fact that B is constant on $\operatorname{zer}(A+B)$ follows from the cocoercivity of B and the monotonicity of A. A proof of this statement when A is the subdifferential of a proper, convex and lower semicontinuous function is given for instance in [1, Lemma 2.7].

Take an arbitrary $x^* \in \operatorname{zer}(A+B)$. From the definition of the resolvent we have for every $t \in [0,+\infty)$

$$-B(x(t)) - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t)) \in A\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t)\right), \tag{5. 58}$$

which combined with $-Bx^* \in Ax^*$ and the monotonicity of A leads to

$$0 \leq \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t) - x^*, -B(x(t)) + Bx^* - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t))\right\rangle. \tag{5. 59}$$

After using the cocoercivity of B we obtain for every $t \in [0,+\infty)$

$$\begin{aligned}\beta\|B(x(t))-Bx^*\|^2 \leq & \left\langle \frac{1}{\lambda(t)}\ddot{x}(t)+\frac{1}{\lambda(t)}\Gamma(\dot{x}(t)), -B(x(t))+Bx^* \right\rangle \\ & - \frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t)+\Gamma(\dot{x}(t))\|^2 \\ & + \left\langle x(t)-x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t)) \right\rangle \\ \leq & \frac{1}{2\beta}\left\| \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) \right\|^2 + \frac{\beta}{2}\|B(x(t))-Bx^*\|^2 \\ & + \left\langle x(t)-x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t)) \right\rangle.\end{aligned}$$

For evaluating the last term of the above inequality we use the functions $h : [0,+\infty) \to \mathbb{R}$,

$$h(t) = \frac{1}{2}\|x(t)-x^*\|^2$$

and $p : [0,+\infty) \to \mathbb{R}$,

$$p(t) = \frac{1}{2}\left\langle (\Gamma - \gamma\,\mathrm{Id})\,(x(t)-x^*), x(t)-x^* \right\rangle,$$

already used in the proof of Theorem 5.10. For every $t \in [0,+\infty)$ we have

$$\langle x(t)-x^*, \ddot{x}(t)\rangle = \ddot{h}(t) - \|\dot{x}(t)\|^2$$

and

$$\dot{p}(t) = \langle x(t)-x^*, \Gamma(\dot{x}(t))\rangle - \gamma\langle x(t)-x^*, \dot{x}(t)\rangle = \langle x(t)-x^*, \Gamma(\dot{x}(t))\rangle - \gamma\dot{h}(t),$$

hence

$$\left\langle x(t)-x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t)) \right\rangle = -\frac{1}{\eta\lambda(t)}\left(\ddot{h}(t)+\gamma\dot{h}(t)+\dot{p}(t)-\|\dot{x}(t)\|^2\right). \tag{5. 60}$$

Consequently, for every $t \in [0,+\infty)$ it holds

$$\begin{aligned}\frac{\beta}{2}\|B(x(t))-Bx^*\|^2 \leq & \frac{1}{2\beta}\left\| \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) \right\|^2 \\ & - \frac{1}{\eta\lambda(t)}\left(\ddot{h}(t)+\gamma\dot{h}(t)+\dot{p}(t)-\|\dot{x}(t)\|^2\right).\end{aligned} \tag{5. 61}$$

By taking into account (A5) we obtain a lower bound $\underline{\lambda}$ such that for every $t \in [0,+\infty)$ one has

$$0 < \underline{\lambda} \leq \lambda(t) \leq \frac{\delta\gamma^2}{2(1+\theta)}.$$

By multiplying (5. 61) with $\lambda(t)$ we obtain for every $t \in [0,+\infty)$ that

$$\frac{\beta\underline{\lambda}}{2}\|B(x(t))-Bx^*\|^2+\frac{1}{\eta}\left(\ddot{h}(t)+\gamma\dot{h}(t)+\dot{p}(t)\right) \leq \frac{1}{2\beta\underline{\lambda}}\|\ddot{x}(t)+\Gamma(\dot{x}(t))\|^2+\frac{1}{\eta}\|\dot{x}(t)\|^2.$$

After integration we obtain that for every $T \in [0,+\infty)$

$$\begin{aligned}& \frac{\beta\underline{\lambda}}{2}\int_0^T \|B(x(t))-Bx^*\|^2 dt + \frac{1}{\eta}\left(\dot{h}(T)-\dot{h}(0)+\gamma h(T)-\gamma h(0)+p(T)-p(0)\right) \\ & \leq \int_0^T \left(\frac{1}{2\beta\underline{\lambda}}\|\ddot{x}(t)+\Gamma(\dot{x}(t))\|^2 + \frac{1}{\eta}\|\dot{x}(t)\|^2\right) dt.\end{aligned}$$

By using that $\dot{x}, \ddot{x} \in L^2([0,+\infty);\mathcal{H})$, $h(T) \geq 0, p(T) \geq 0$ for every $T \in [0,+\infty)$ and $\lim_{T\to+\infty} \dot{h}(T) = 0$, it follows that $B(x(\cdot)) - Bx^* \in L^2([0,+\infty);\mathcal{H})$.

Further, by taking into consideration Remark 5.1(b), we have

$$\begin{aligned}\frac{d}{dt}\left(\frac{1}{2}\|B(x(t)) - Bx^*\|^2\right) &= \left\langle B(x(t)) - Bx^*, \frac{d}{dt}(Bx(t))\right\rangle \\ &\leq \frac{1}{2}\|B(x(t)) - Bx^*\|^2 + \frac{1}{2\beta^2}\|\dot{x}(t)\|^2\end{aligned}$$

and from here, in light of Lemma 5.2, it follows that $\lim_{t\to+\infty} B(x(t)) = Bx^*$.

(v) Let x^* be the unique element of $\operatorname{zer}(A+B)$. For the beginning we suppose that A is uniformly monotone with corresponding function $\phi_A : [0,+\infty) \to [0,+\infty]$, which is increasing and vanishes only at 0.

By similar arguments as in the proof of statement (iv), for every $t \in [0,+\infty)$ we have

$$\begin{aligned}&\phi_A\left(\left\|\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t) - x^*\right\|\right) \leq \\ &\left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t) - x^*, -B(x(t)) + Bx^* - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t))\right\rangle,\end{aligned}$$

which combined with the inequality

$$\langle x(t) - x^*, B(x(t)) - Bx^*\rangle \geq 0$$

yields

$$\begin{aligned}&\phi_A\left(\left\|\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t) - x^*\right\|\right) \\ &\leq \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)), -B(x(t)) + Bx^*\right\rangle - \frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t) + \Gamma(\dot{x}(t))\|^2 \\ &\quad + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t))\right\rangle \\ &\leq \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)), -B(x(t)) + Bx^*\right\rangle \\ &\quad + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t))\right\rangle.\end{aligned}$$

As λ is bounded by positive constants, by using (i)-(iv) it follows that the right-hand side of the last inequality converges to 0 as $t \to +\infty$. Hence

$$\lim_{t\to+\infty} \phi_A\left(\left\|\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t) - x^*\right\|\right) = 0$$

and the properties of the function ϕ_A allow to conclude that $\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t) - x^*$ converges strongly to 0 as $t \to +\infty$. By using again the boundedness of λ and (ii) we obtain that $x(t)$ converges strongly to x^* as $t \to +\infty$.

Finally, suppose that B is uniformly monotone with corresponding function $\phi_B : [0,+\infty) \to [0,+\infty]$, which is increasing and vanishes only at 0. The conclusion follows by letting t in the inequality

$$\langle x(t) - x^*, B(x(t)) - Bx^*\rangle \geq \phi_B(\|x(t) - x^*\|) \ \forall t \in [0,+\infty)$$

converge to $+\infty$ and by using that x is bounded and $\lim_{t\to+\infty}(B(x(t) - Bx^*) = 0$.
□

Remark 5.13 We would like to emphasize the fact that the statements in Theorem 5.11 remain valid also for $\eta := 2\beta$. Indeed, in this case the cocoercivity of B implies that $\operatorname{Id} - \eta B$ is nonexpansive, hence the operator $R = J_{\eta A} \circ (\operatorname{Id} - \eta B)$ used in the proof is nonexpansive, too, and so the statements in (i)-(iii) follow from Corollary 5.2. Furthermore, the proof of the statements (iv) and (v) can be repeated also for $\eta = 2\beta$.

In the remaining of this section we turn our attention to optimization problems of the form

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

where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ is a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$.

In the following statement we approach the minimizers of $f + g$ via the second order dynamical system

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t) \left[x(t) - \operatorname{prox}_{\eta f}\left(x(t) - \eta \nabla g(x(t))\right)\right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{5. 62}$$

Corollary 5.4 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ by a proper, convex and lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$ such that $\operatorname{argmin}_{x \in \mathcal{H}}\{f(x)+g(x)\} \neq \emptyset$. Let $\eta \in (0, 2\beta]$ and set $\delta := (4\beta - \eta)/(2\beta)$. Let $\Gamma : \mathcal{H} \to \mathcal{H}$ be an operator fulfilling (A1), $\lambda : [0, +\infty) \to (0, +\infty)$ be a function fulfilling (A5), $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of (5. 62). Then the following statements are true:*

(i) $x(\cdot)$ *is bounded and* $\dot{x}, \ddot{x}, \big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta \nabla g)\big)x \in L^2([0, +\infty); \mathcal{H})$*;*

(ii) $\lim_{t \to +\infty} \dot{x}(t) = \lim_{t \to +\infty} \ddot{x}(t) = \lim_{t \to +\infty} \big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta \nabla g)\big)(x(t)) = 0$*;*

(iii) $x(t)$ *converges weakly to a minimizer of* $f + g$ *as* $t \to +\infty$*;*

(iv) if x^* *is a minimizer of* $f + g$*, then* $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t \to +\infty} \nabla g(x(t)) = \nabla g(x^*)$ *and* ∇g *is constant on* $\operatorname{argmin}_{x \in \mathcal{H}}\{f(x)+g(x)\}$*;*

(v) if f *or* g *is uniformly convex, then* $x(t)$ *converges strongly to the unique minimizer of* $f + g$ *as* $t \to +\infty$*.*

Proof. The statements are direct consequences of the corresponding ones from Theorem 5.11 (see also Remark 5.13), by choosing $A := \partial f$ and $B := \nabla g$, by taking into account that

$$\operatorname{zer}(\partial f + \nabla g) = \operatorname*{argmin}_{x \in \mathcal{H}}\{f(x) + g(x)\}$$

and by making use of the Baillon-Haddad Theorem, which says that ∇g is $1/\beta$-Lipschitz continuous if and only if ∇g is β-cocoercive (see [26, Corollary 18.16]). For statement (v) we also use the fact that if f is uniformly convex with modulus ϕ, then ∂f is uniformly monotone with modulus 2ϕ (see [26, Example 22.3(iii)]). □

Remark 5.14 Consider again the setting in Remark 5.12, namely, when $\Gamma = \gamma \operatorname{Id}$ for $\gamma > 0$ and $\lambda(t) = 1$ for every $t \in [0, +\infty)$. Furthermore, for C a nonempty, convex, closed subset of $\mathcal{H}$, let $f = \delta_C$ be the indicator function of C, which is defined as being equal to 0 for $x \in C$ and to $+\infty$, else. The dynamical system (5. 62) attached in this setting to the minimization of g over C becomes

$$\left\{ \begin{array}{l} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) - P_C\big(x(t) - \eta\nabla g(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{array} \right. \tag{5. 63}$$

where P_C denotes the projection onto the set C.

The convergence of the trajectories of (5. 63) has been studied in [8, Theorem 3.1] under the conditions $\gamma^2 > 2$ and $0 < \eta \leq 2\beta$. In this case assumption (A5) trivially holds by choosing θ such that $0 < \theta \leq (\gamma^2 - 2)/2 \leq (\delta\gamma^2 - 2)/2$. Thus, in order to verify (A5) in case $\lambda(t) = 1$ for every $t \in [0, +\infty)$ one needs to equivalently assume that $\gamma^2 > 2/\delta$. Since $\delta \geq 1$, this provides a slight improvement over [8, Theorem 3.1] in what concerns the choice of γ. We refer the reader also to [7] for an analysis of the convergence rates of trajectories of the dynamical system (5. 63) when g is endowed with supplementary properties.

For the two main convergence statements provided in this section it was essential to choose the step size η in the interval $(0, 2\beta]$ (see Theorem 5.11, Remark 5.13 and Corollary 5.4). This, because of the fact that in this way we were able to guarantee for the generated trajectories the existence of the limit $\lim_{t\to+\infty} \|x(t) - x^*\|^2$, where x^* denotes a solution of the problem under investigation. It is interesting to observe that, when dealing with convex optimization problems, one can go also beyond this classical restriction concerning the choice of the step size (a similar phenomenon has been reported also in [1, Section 5.2]). This is pointed out in the following result, which is valid under the assumption

(A6) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous and there exist $a, \theta, \theta' > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\lambda}(t) \geq 0 \text{ and } \frac{1}{\beta}\left(\theta' + \frac{a}{2}\|\Gamma - \gamma \operatorname{Id}\|\right) \leq \lambda(t) \leq \frac{\gamma^2}{\eta\theta + \frac{\eta}{\beta} + \frac{\eta}{2a}\|\Gamma - \gamma \operatorname{Id}\| + 1}, \tag{5. 64}$$

and for the proof of which we use instead of $\|x(\cdot) - x^*\|^2$ a modified energy functional.

Corollary 5.5 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ by a proper, convex and lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$ such that* $\operatorname{argmin}_{x\in\mathcal{H}}\{f(x) + g(x)\} \neq \emptyset$. *Let be $\eta > 0$, $\Gamma : \mathcal{H} \to \mathcal{H}$ be an operator fulfilling* (A1), *$\lambda : [0, +\infty) \to (0, +\infty)$ be a function fulfilling* (A6), *$u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 62). *Then the following statements are true:*

(i) *$x(\cdot)$ is bounded and $\dot{x}, \ddot{x}, \big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)x \in L^2([0, +\infty); \mathcal{H})$;*

(ii) *$\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} \big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)(x(t)) = 0$;*

(iii) *$x(t)$ converges weakly to a minimizer of $f + g$ as $t \to +\infty$;*

(iv) *if x^* is a minimizer of $f + g$, then $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t\to+\infty} \nabla g(x(t)) = \nabla g(x^*)$ and ∇g is constant on* $\operatorname{argmin}_{x\in\mathcal{H}}\{f(x) + g(x)\}$;

(v) *if f or g is uniformly convex, then $x(t)$ converges strongly to the unique minimizer of $f + g$ as $t \to +\infty$.*

Proof. Consider an arbitrary element $x^* \in \operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\} = \operatorname{zer}(\partial f + \nabla g)$. Similarly to the proof of Theorem 5.11(iv), we derive for every $t \in [0,+\infty)$ (see the first inequality after (5. 59))

$$\begin{aligned}
&\beta\|\nabla g(x(t)) - \nabla g(x^*)\|^2 \\
&\quad \le \frac{1}{\lambda(t)}\Big(\langle \ddot{x}(t), -\nabla g(x(t)) + \nabla g(x^*)\rangle + \langle \Gamma(\dot{x}(t)), -\nabla g(x(t)) + \nabla g(x^*)\rangle \Big) \\
&\qquad - \frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t) + \Gamma(\dot{x}(t))\|^2 + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t)) \right\rangle .
\end{aligned} \tag{5. 65}$$

In what follows we evaluate the right-hand side of the above inequality and introduce to this end the function

$$q : [0,+\infty) \to \mathbb{R},\ q(t) = g(x(t)) - g(x^*) - \langle \nabla g(x^*), x(t) - x^* \rangle .$$

Due to the convexity of g one has

$$q(t) \ge 0 \ \forall t \ge 0.$$

Further, for every $t \in [0,+\infty)$

$$\dot{q}(t) = \langle \dot{x}(t), \nabla g(x(t)) - \nabla g(x^*)\rangle ,$$

thus

$$\begin{aligned}
&\langle \Gamma(\dot{x}(t)), -\nabla g(x(t)) + \nabla g(x^*)\rangle \\
&\quad = -\gamma\dot{q}(t) + \langle (\Gamma - \gamma \operatorname{Id})(\dot{x}(t)), -\nabla g(x(t)) + \nabla g(x^*)\rangle \\
&\quad \le -\gamma\dot{q}(t) + \frac{1}{2a}\|\Gamma - \gamma \operatorname{Id}\|\|\dot{x}(t)\|^2 + \frac{a}{2}\|\Gamma - \gamma \operatorname{Id}\|\|\nabla g(x(t)) - \nabla g(x^*)\|^2.
\end{aligned} \tag{5. 66}$$

On the other hand, for every $t \in [0,+\infty)$

$$\ddot{q}(t) = \langle \ddot{x}(t), \nabla g(x(t)) - \nabla g(x^*)\rangle + \left\langle \dot{x}(t), \frac{d}{dt}\nabla g(x(t)) \right\rangle ,$$

hence

$$\langle \ddot{x}(t), -\nabla g(x(t)) + \nabla g(x^*)\rangle \le -\ddot{q}(t) + \frac{1}{\beta}\|\dot{x}(t)\|^2. \tag{5. 67}$$

Further, we have for every $t \in [0,+\infty)$ (see also (5. 42) and (5. 41))

$$\begin{aligned}
\frac{1}{\lambda(t)}\|\ddot{x}(t) + \Gamma(\dot{x}(t))\|^2 &= \frac{1}{\lambda(t)}\|\ddot{x}(t)\|^2 + \frac{1}{\lambda(t)}\frac{d}{dt}\big(\langle \dot{x}(t), \Gamma(\dot{x}(t))\rangle \big) + \frac{1}{\lambda(t)}\|\Gamma(\dot{x}(t))\|^2 \\
&\ge \frac{1}{\lambda(t)}\|\ddot{x}(t)\|^2 + \frac{d}{dt}\left(\frac{1}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t))\rangle \right) \\
&\quad + \gamma\frac{\dot{\lambda}(t)}{\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{\gamma^2}{\lambda(t)}\|\dot{x}(t)\|^2.
\end{aligned} \tag{5. 68}$$

Finally, by multiplying (5. 65) with $\lambda(t)$ and by using (5. 66), (5. 67), (5. 68) and

(5. 60) we obtain after rearranging the terms for every $t \in [0, +\infty)$ that

$$\begin{aligned}
&\left(\beta\lambda(t) - \frac{a}{2}\|\Gamma - \gamma \operatorname{Id}\|\right) \|\nabla g(x(t)) - \nabla g(x^*)\|^2 \\
&+ \frac{d^2}{dt^2}\left(\frac{1}{\eta}h + q\right) + \gamma \frac{d}{dt}\left(\frac{1}{\eta}h + q\right) \\
&+ \ \frac{1}{\eta}\dot{p}(t) + \frac{1}{\eta}\frac{d}{dt}\left(\frac{1}{\lambda(t)}\left\langle \dot{x}(t), \Gamma(\dot{x}(t))\right\rangle\right) \\
&+ \ \left(\frac{\gamma^2}{\eta\lambda(t)} + \frac{\gamma\dot{\lambda}(t)}{\eta\lambda^2(t)} - \frac{1}{\beta} - \frac{1}{\eta} - \frac{1}{2a}\|\Gamma - \gamma \operatorname{Id}\|\right)\|\dot{x}(t)\|^2 \\
&+ \frac{1}{\eta\lambda(t)}\|\ddot{x}(t)\|^2 \\
&\leq 0.
\end{aligned}$$

and, further, via (A6)

$$\begin{aligned}
&\theta'\|\nabla g(x(t)) - \nabla g(x^*)\|^2 + \frac{d^2}{dt^2}\left(\frac{1}{\eta}h + q\right) + \gamma\frac{d}{dt}\left(\frac{1}{\eta}h + q\right) + \frac{1}{\eta}\dot{p}(t) \\
&+ \frac{1}{\eta}\frac{d}{dt}\left(\frac{1}{\lambda(t)}\left\langle \dot{x}(t), \Gamma(\dot{x}(t))\right\rangle\right) + \theta\|\dot{x}(t)\|^2 + \frac{1}{\eta\lambda(t)}\|\ddot{x}(t)\|^2 \\
&\leq 0.
\end{aligned} \tag{5. 69}$$

This implies that the function

$$t \mapsto \frac{d}{dt}\left(\frac{1}{\eta}h + q\right)(t) + \gamma\left(\frac{1}{\eta}h + q\right)(t) + \frac{1}{\eta}p(t) + \frac{1}{\eta}\left(\frac{1}{\lambda(t)}\left\langle \dot{x}(t), \Gamma(\dot{x}(t))\right\rangle\right) \tag{5. 70}$$

is monotonically decreasing. Arguing as in the proof of Theorem 5.10, by taking into account that λ has positive upper and lower bounds, it follows that

$$\frac{1}{\eta}h + q, h, q, x, \dot{x}, \dot{h}, \dot{q} \text{ are bounded,}$$

$\dot{x}, \ddot{x}$ and $\big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)x \in L^2([0,+\infty); \mathcal{H})$ and $\lim_{t\to+\infty}\dot{x}(t) = 0$. Since $\frac{d}{dt}\big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)x \in L^2([0,+\infty); \mathcal{H})$ (see Remark 5.1(b)), we derive from Lemma 5.2 that $\lim_{t\to+\infty}\big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)(x(t)) = 0$. As $\ddot{x}(t) = -\Gamma(\dot{x}(t)) - \lambda(t)\big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)(x(t))$ for every $t \in [0, +\infty)$, we obtain that $\lim_{t\to+\infty}\ddot{x}(t) = 0$. From (5. 69) it also follows that $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0,+\infty); \mathcal{H})$ and, by applying again Lemma 5.2, it yields $\lim_{t\to+\infty}\nabla g(x(t)) = \nabla g(x^*)$. In this way the statements (i), (ii) and (iv) are shown.

(iii) Since the function in (5. 70) is monotonically decreasing, from (i), (ii) and (iv) it follows that the limit $\lim_{t\to+\infty}\left(\gamma\left(\frac{1}{\eta}h + q\right)(t) + \frac{1}{\eta}p(t)\right)$ exists and it is a real number.

Consider again the renorming of the space already used in the proof of Theorem 5.10(iii). As x^* has been chosen as an arbitrary minimizer of $f + g$ and taking into account the definition of p and the new norm, we conclude that for all $x^* \in \operatorname{argmin}_{x\in\mathcal{H}}\{f(x) + g(x)\}$ the limit $\lim_{t\to+\infty} E(t, x^*) \in \mathbb{R}$, exists, where

$$E(t, x^*) = \frac{1}{2\eta}|||x(t) - x^*|||^2 + g(x(t)) - g(x^*) - \left\langle \nabla g(x^*), x(t) - x^*\right\rangle.$$

In what follows we use a similar technique as in [31] (see, also, [1, Section 5.2]). Since $x(\cdot)$ is bounded, it has at least one weak sequential cluster point.

We prove first that each weak sequential cluster point of $x(\cdot)$ is a minimizer of $f+g$. Let $x^* \in \operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\}$ and $t_n \to +\infty$ (as $n \to +\infty$) be such that $(x(t_n))_{n\in\mathbb{N}}$ converges weakly to $\overline{x}$. Since $(x(t_n), \nabla g(x(t_n))) \in \operatorname{gr}(\nabla g)$, $\lim_{n\to+\infty} \nabla g(x(t_n)) = \nabla g(x^*)$ and $\operatorname{gr}(\nabla g)$ is sequentially closed in the weak-strong topology, we obtain $\nabla g(\overline{x}) = \nabla g(x^*)$.

From (5. 58) written for $t = t_n$, $A = \partial f$ and $B = \nabla g$, by letting n converge to $+\infty$ and by using that $\operatorname{gr}(\partial f)$ is sequentially closed in the weak-strong topology, we obtain $-\nabla g(x^*) \in \partial f(\overline{x})$. This, combined with $\nabla g(\overline{x}) = \nabla g(x^*)$ delivers $-\nabla g(\overline{x}) \in \partial f(\overline{x})$, hence $\overline{x} \in \operatorname{zer}(\partial f + \nabla g) = \operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\}$.

Next we show that $x(\cdot)$ has at most one weak sequential cluster point, which will actually guarantee that it has exactly one weak sequential cluster point. This will imply the weak convergence of the trajectory to a minimizer of $f+g$.

Let x_1^*, x_2^* be two weak sequential cluster points of $x(\cdot)$. This means that there exist $t_n \to +\infty$ (as $n \to +\infty$) and $t_n' \to +\infty$ (as $n \to +\infty$) such that $(x(t_n))_{n\in\mathbb{N}}$ converges weakly to x_1^* (as $n \to +\infty$) and $(x(t_n'))_{n\in\mathbb{N}}$ converges weakly to x_2^* (as $n \to +\infty$). Since $x_1^*, x_2^* \in \operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\}$, we have $\lim_{t\to+\infty} E(t, x_1^*) \in \mathbb{R}$ and $\lim_{t\to+\infty} E(t, x_2^*) \in \mathbb{R}$, hence $\exists \lim_{t\to+\infty}(E(t,x_1^*) - E(t,x_2^*)) \in \mathbb{R}$. We obtain

$$\exists \lim_{t\to+\infty}\left(\frac{1}{\eta}\langle\langle x(t), x_2^* - x_1^*\rangle\rangle + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x(t)\rangle\right) \in \mathbb{R},$$

which, when expressed by means of the sequences $(t_n)_{n\in\mathbb{N}}$ and $(t_n')_{n\in\mathbb{N}}$, leads to $\frac{1}{\eta}\langle\langle x_1^*, x_2^* - x_1^*\rangle\rangle + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x_1^*\rangle = \frac{1}{\eta}\langle\langle x_2^*, x_2^* - x_1^*\rangle\rangle + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x_2^*\rangle$. This is the same with

$$\frac{1}{\eta}|||x_1^* - x_2^*|||^2 + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x_2^* - x_1^*\rangle = 0$$

and by the monotonicity of ∇g we conclude that $x_1^* = x_2^*$.

(v) The proof of this statement follows in analogy to the one of the corresponding statement of Theorem 5.11(v) written for $A = \partial f$ and $B = \nabla g$. $\square$

Remark 5.15 When $\Gamma = \gamma \operatorname{Id}$ for $\gamma > 0$, in order to verify the left-hand side of the second statement in assumption (A6) one can take $\theta' := \beta \inf_{t\geq 0} \lambda(t)$. Thus, (5. 64) amounts in this case to the existence of $\theta > 0$ such that

$$\lambda(t) \leq \frac{\gamma^2}{\eta\theta + \frac{\eta}{\beta} + 1}.$$

When one takes $\lambda(t) = 1$ for every $t \in [0, +\infty)$, this is verified if and only if $\gamma^2 > \frac{\eta}{\beta} + 1$. In other words, (A6) allows in this particular setting a more relaxed choice for the parameters γ, η and β, beyond the standard assumptions $0 < \eta \leq 2\beta$ and $\gamma^2 > 2$ considered in [8].

In the following we provide a rate for the convergence of a convex and (Fréchet) differentiable function with Lipschitz continuous gradient $g : \mathcal{H} \to \mathbb{R}$ along the ergodic trajectory generated by

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\nabla g(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0 \end{cases} \tag{5. 71}$$

to the minimum value of g. To this end we make the following assumption

(A7) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous and there exists $\zeta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$0 < \zeta \leq \gamma\lambda(t) - \dot{\lambda}(t). \tag{5. 72}$$

Let us mention that the following result is in the spirit of a convergence rate statement given for the objective function values on a sequence iteratively generated by an inertial-type algorithm recently obtained in [89, Theorem 1].

Theorem 5.12 *Let $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$ such that* $\operatorname{argmin}_{x\in\mathcal{H}} g(x) \neq \emptyset$. *Let $\Gamma : \mathcal{H} \to \mathcal{H}$ be an operator fulfilling* (A1)*, $\lambda : [0,+\infty) \to (0,+\infty)$ a function fulfilling* (A7)*, $u_0, v_0 \in \mathcal{H}$ and $x : [0,+\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 71).

Then for every minimizer x^* of g and every $T > 0$ it holds

$$\begin{aligned}
0 &\leq g\left(\frac{1}{T}\int_0^T x(t)dt\right) - g(x^*)\\
&\leq \frac{1}{2\zeta T}\left[\|v_0 + \gamma(u_0 - x^*)\|^2 + \left(\gamma\|\Gamma - \gamma\operatorname{Id}\| + \frac{\lambda(0)}{\beta}\right)\|u_0 - x^*\|^2\right].
\end{aligned}$$

Proof. The existence and uniqueness of the trajectory of (5. 71) follow from Theorem 5.9. Let be $x^* \in \operatorname{argmin}_{x\in\mathcal{H}} g(x)$, $T > 0$ and consider again the function $p : [0,+\infty) \to \mathbb{R}$,

$$p(t) = \frac{1}{2}\left\langle (\Gamma - \gamma\operatorname{Id})\,(x(t) - x^*), x(t) - x^*\right\rangle$$

which we defined in (5. 38). By using (5. 71), the formula for the derivative of p, the positive semidefinitness of $\Gamma - \gamma\operatorname{Id}$, the convexity of g and (A7) we get for almost every $t \in [0,+\infty)$

$$\begin{aligned}
&\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t) + \gamma(x(t) - x^*)\|^2 + \gamma p(t) + \lambda(t)g(x(t))\right)\\
&= \langle \ddot{x}(t) + \gamma\dot{x}(t), \dot{x}(t) + \gamma(x(t) - x^*)\rangle + \gamma\left\langle (\Gamma - \gamma\operatorname{Id})(\dot{x}(t)), x(t) - x^*\right\rangle\\
&\quad + \dot{\lambda}(t)g(x(t)) + \lambda(t)\left\langle \dot{x}(t), \nabla g(x(t))\right\rangle\\
&= \langle -(\Gamma - \gamma\operatorname{Id})(\dot{x}(t)) - \lambda(t)\nabla g(x(t)), \dot{x}(t) + \gamma(x(t) - x^*)\rangle\\
&\quad + \langle (\Gamma - \gamma\operatorname{Id})(\dot{x}(t)), \gamma(x(t) - x^*)\rangle + \dot{\lambda}(t)g(x(t)) + \lambda(t)\left\langle \dot{x}(t), \nabla g(x(t))\right\rangle\\
&\leq -\gamma\lambda(t)\left\langle \nabla g(x(t)), x(t) - x^*\right\rangle + \dot{\lambda}(t)g(x(t))\\
&\leq (\dot{\lambda}(t) - \gamma\lambda(t))(g(x(t)) - g(x^*)) + \dot{\lambda}(t)g(x^*)\\
&\leq -\zeta(g(x(t)) - g(x^*)) + \dot{\lambda}(t)g(x^*).
\end{aligned}$$

We obtain after integration

$$\begin{aligned}
&\frac{1}{2}\|\dot{x}(T) + \gamma(x(T) - x^*)\|^2 + \gamma p(T) + \lambda(T)g(x(T))\\
&- \left(\frac{1}{2}\|\dot{x}(0) + \gamma(x(0) - x^*)\|^2 + \gamma p(0) + \lambda(0)g(x(0))\right)\\
&+ \zeta\int_0^T (g(x(t)) - g(x^*))dt\\
&\leq (\lambda(T) - \lambda(0))g(x^*).
\end{aligned}$$

Be neglecting the nonnegative terms on the left-hand side of this inequality and by using that $g(x(T)) \geq g(x^*)$, it yields

$$\zeta\int_0^T (g(x(t)) - g(x^*))dt \leq \frac{1}{2}\|v_0 + \gamma(u_0 - x^*)\|^2 + \gamma p(0) + \lambda(0)(g(u_0) - g(x^*)).$$

The conclusion follows by using

$$\begin{aligned}p(0) &= \frac{1}{2}\left\langle (\Gamma - \gamma\,\mathrm{Id})(u_0 - x^*), u_0 - x^*\right\rangle \\ &\leq \frac{1}{2}\|\Gamma - \gamma\,\mathrm{Id}\,\|\|u_0 - x^*\|^2,\end{aligned}$$

and

$$g(u_0) - g(x^*) \leq \frac{1}{2\beta}\|u_0 - x^*\|^2,$$

which is a consequence of the descent lemma (see Lemma 1.4 and notice that $\nabla g(x^*) = 0$), and the inequality

$$g\left(\frac{1}{T}\int_0^T x(t)dt\right) - g(x^*) \leq \frac{1}{T}\int_0^T (g(x(t)) - g(x^*))dt,$$

which holds since g is convex. □

Remark 5.16 Under assumption (A7) on the relaxation function λ, we obtain in the above theorem (only) the convergence of the function g along the ergodic trajectory to a global minimum value. If one is interested also in the (weak) convergence of the trajectory to a minimizer of g, this follows via Theorem 5.10 when λ is assumed to fulfill (A2) (notice that if x converges weakly to a minimizer of g, then from the Cesaro-Stolz Theorem one also obtains the weak convergence of the ergodic trajectory $T \mapsto \frac{1}{T}\int_0^T x(t)dt$ to the same minimizer).

Take $a \geq 0$, $b > 1/(\beta\gamma^2)$ and $0 \leq \rho \leq \gamma$. Then

$$\lambda(t) = \frac{1}{ae^{-\rho t} + b}$$

is an example of a relaxation function which verifies assumption (A2) (with $0 < \theta \leq b\beta\gamma^2 - 1$) and assumption (A7) (with $0 < \zeta \leq \gamma b/(a+b)^2$).

5.2.3 Variable damping parameters

In this section we carry out a similar analysis as in the previous subsection, however, for second order dynamical systems having as damping coefficient a function depending on time. We refer the reader to [13,65,66,126] for other works where second order differential equations with time dependent damping have been considered and investigated in connection with optimization problems.

As starting point for our investigations we consider the dynamical system

$$\left\{\begin{array}{l} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)B(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{array}\right. \tag{5. 73}$$

where $B : \mathcal{H} \to \mathcal{H}$ is a cocoercive operator, $\lambda, \gamma : [0, +\infty) \to [0, +\infty)$ are Lebesgue measurable functions and $u_0, v_0 \in \mathcal{H}$.

The existence and uniqueness of strong global solutions of (5. 73) can be shown by using the same techniques as in the proof of Theorem 5.9, provided that $\lambda(\cdot), \gamma(\cdot) \in L^1_{\text{loc}}([0, +\infty))$. For the convergence of the trajectories we need the following assumption

(A2') $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq \frac{1+\theta}{\beta}. \tag{5. 74}$$

According to Definition 5.1 and Remark 5.1(a), $\dot{\lambda}(t), \dot{\gamma}(t)$ exist for almost almost every $t \in [0,+\infty)$ and $\dot{\lambda}, \dot{\gamma}$ are Lebesgue integrable on each interval $[0,b]$ for $0 < b < +\infty$. This combined with $\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t)$, yields the existence of a positive lower bound for λ and for a positive upper bound for γ. Using further the second assumption in (5. 74) provides also a positive upper bound for λ and a positive lower bound for γ. The couple of functions

$$\lambda(t) = \frac{1}{ae^{-\rho t}+b} \text{ and } \gamma(t) = a'e^{-\rho' t} + b',$$

where $a, a', \rho, \rho' \geq 0$ and $b, b' > 0$ fulfill the inequality $b'^2 b > 1/\beta$, verify the conditions in assumption (A2').

We state now the convergence result.

Theorem 5.13 *Let $B : \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator for $\beta > 0$ such that* $\operatorname{zer} B \neq \emptyset$, *$\lambda, \gamma : [0,+\infty) \to (0,+\infty)$ be functions fulfilling* (A2') *and $u_0, v_0 \in \mathcal{H}$. Let $x : [0,+\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 73)*. Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, Bx \in L^2([0,+\infty);\mathcal{H})$;

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} B(x(t)) = 0$*;*

(iii) $x(t)$ converges weakly to an element in $\operatorname{zer} B$ *as $t \to +\infty$.*

Proof. With the notations in the proof of Theorem 5.10 and by appealing to similar arguments one obtains for every $t \in [0,+\infty)$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \frac{\beta}{\lambda(t)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 \leq \|\dot{x}(t)\|^2$$

or, equivalently,

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \frac{\beta\gamma(t)}{\lambda(t)}\frac{d}{dt}\left(\|\dot{x}(t)\|^2\right) + \left(\frac{\beta\gamma^2(t)}{\lambda(t)} - 1\right)||\dot{x}(t)||^2 + \frac{\beta}{\lambda(t)}||\ddot{x}(t)||^2 \leq 0.$$

Combining this inequality with

$$\frac{\gamma(t)}{\lambda(t)}\frac{d}{dt}\left(\|\dot{x}(t)\|^2\right) = \frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right) - \frac{\dot{\gamma}(t)\lambda(t) - \gamma(t)\dot{\lambda}(t)}{\lambda^2(t)}\|\dot{x}(t)\|^2$$

and

$$\gamma(t)\dot{h}(t) = \frac{d}{dt}(\gamma h)(t) - \dot{\gamma}(t)h(t) \geq \frac{d}{dt}(\gamma h)(t), \tag{5. 75}$$

it yields for every $t \in [0,+\infty)$

$$\begin{aligned}&\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \beta\frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right)\\&+\left(\frac{\beta\gamma^2(t)}{\lambda(t)} + \beta\frac{-\dot{\gamma}(t)\lambda(t) + \gamma(t)\dot{\lambda}(t)}{\lambda^2(t)} - 1\right)||\dot{x}(t)||^2 + \frac{\beta}{\lambda(t)}||\ddot{x}(t)||^2\\&\leq 0.\end{aligned}$$

Now, assumption (A2') delivers for almost every $t \in [0,+\infty)$ the inequality

$$\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \beta\frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right) + \theta||\dot{x}(t)||^2 + \frac{\beta}{\lambda(t)}||\ddot{x}(t)||^2 \leq 0.$$

This implies that the function $t \mapsto \dot{h}(t) + \gamma(t)h(t) + \beta\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2$ is monotonically decreasing and from here one obtains the conclusion following the lines of the proof of Theorem 5.10, by taking also into account that $\exists \lim_{t\to+\infty}\gamma(t) \in (0,+\infty)$. □

When $T : \mathcal{H} \to \mathcal{H}$ is a nonexpansive operator one obtains for the dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\big(x(t) - T(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0 \end{cases} \tag{5. 76}$$

and by making the assumption

(A3') $\lambda, \gamma : [0,+\infty) \to (0,+\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0,+\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq 2(1+\theta) \tag{5. 77}$$

the following result which can been seen as a counterpart to Corollary 5.2.

Corollary 5.6 *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator such that* $\operatorname{Fix} T \neq \emptyset$, *$\lambda, \gamma : [0,+\infty) \to (0,+\infty)$ be functions fulfilling* (A3') *and $u_0, v_0 \in \mathcal{H}$. Let $x : [0,+\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 76)*. Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, (\operatorname{Id} - T)x \in L^2([0,+\infty);\mathcal{H})$;

(ii) $\lim_{t\to+\infty}\dot{x}(t) = \lim_{t\to+\infty}\ddot{x}(t) = \lim_{t\to+\infty}(\operatorname{Id} - T)(x(t)) = 0$*;*

(iii) $x(t)$ converges weakly to a point in $\operatorname{Fix} T$ *as $t \to +\infty$.*

When $R : \mathcal{H} \to \mathcal{H}$ is an α-averaged operator for $\alpha \in (0,1)$ one obtains for the dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\big(x(t) - R(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{5. 78}$$

and by making the assumption

(A4') $\lambda, \gamma : [0,+\infty) \to (0,+\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0,+\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq 2\alpha(1+\theta) \tag{5. 79}$$

the following result which can been seen as a counterpart to Corollary 5.3.

Corollary 5.7 *Let $R : \mathcal{H} \to \mathcal{H}$ be an α-averaged operator for $\alpha \in (0,1)$ such that* $\operatorname{Fix} R \neq \emptyset$*, $\lambda, \gamma : [0,+\infty) \to (0,+\infty)$ be functions fulfilling* (A4') *and $u_0, v_0 \in \mathcal{H}$. Let $x : [0,+\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 78)*. Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, (\operatorname{Id} - R)x \in L^2([0,+\infty);\mathcal{H})$;

(ii) $\lim_{t\to+\infty}\dot{x}(t) = \lim_{t\to+\infty}\ddot{x}(t) = \lim_{t\to+\infty}(\operatorname{Id} - R)(x(t)) = 0$*;*

(iii) $x(t)$ converges weakly to a point in $\operatorname{Fix} R$ *as $t \to +\infty$.*

We come now to the monotone inclusion problem

$$\text{find } 0 \in A(x) + B(x),$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ is a β-cocoercive operator for $\beta > 0$ and assign to it the second order dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\Big[x(t) - J_{\eta A}\Big(x(t) - \eta B(x(t))\Big)\Big] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{5. 80}$$

and make the assumption

(A5') $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq \frac{2(1+\theta)}{\delta}. \tag{5. 81}$$

Theorem 5.14 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ be β-cocoercive operator for $\beta > 0$ such that $\operatorname{zer}(A + B) \neq \emptyset$. Let $\eta \in (0, 2\beta)$ and set $\delta := (4\beta - \eta)/(2\beta)$. Let $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ be functions fulfilling* (A5'), *$u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 80). *Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, \big(\operatorname{Id} - J_{\eta A} \circ (\operatorname{Id} - \eta B)\big)x \in L^2([0, +\infty); \mathcal{H})$;

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} \big(\operatorname{Id} - J_{\eta A} \circ (\operatorname{Id} - \eta B)\big)(x(t)) = 0$;

(iii) $x(t)$ converges weakly to a point in $\operatorname{zer}(A + B)$ as $t \to +\infty$;

(iv) if $x^* \in \operatorname{zer}(A+B)$, then $B(x(\cdot)) - Bx^* \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t\to+\infty} B(x(t)) = Bx^*$ and B is constant on $\operatorname{zer}(A + B)$;

(v) if A or B is uniformly monotone, then $x(t)$ converges strongly to the unique point in $\operatorname{zer}(A + B)$ as $t \to +\infty$.

Proof. The statements (i)-(iii) follow by using the same arguments as in the proof of Theorem 5.11.

(iv) We use again the notations in the proof of Theorem 5.10. Let be an arbitrary $x^* \in \operatorname{zer}(A + B)$. From the definition of the resolvent we have for every $t \in [0, +\infty)$

$$-B(x(t)) - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \in A\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right), \tag{5. 82}$$

which combined with $-Bx^* \in Ax^*$ and the monotonicity of A leads to

$$0 \leq \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*, -B(x(t)) + Bx^* - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t)\right\rangle. \tag{5. 83}$$

The cocoercivity of B yields for every $t \in [0, +\infty)$

$$\begin{aligned}\beta\|B(x(t)) - Bx^*\|^2 \leq & \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t), -B(x(t)) + Bx^*\right\rangle \\ & - \frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 \\ & + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t)\right\rangle \\ \leq & \frac{1}{2\beta}\left\|\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t)\right\|^2 + \frac{\beta}{2}\|B(x(t)) - Bx^*\|^2 \\ & + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t)\right\rangle.\end{aligned}$$

From

$$\left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle = -\frac{1}{\eta\lambda(t)}\left(\ddot{h}(t) + \gamma(t)\dot{h}(t) - \|\dot{x}(t)\|^2\right) \tag{5. 84}$$

we obtain for every $t \in [0, +\infty)$

$$\frac{\beta\lambda(t)}{2}\|B(x(t)) - Bx^*\|^2 + \frac{1}{\eta}\left(\ddot{h}(t) + \gamma(t)\dot{h}(t)\right) \leq \frac{1}{2\beta\lambda(t)}\|\ddot{x}(t) + \gamma(t)(\dot{x}(t))\|^2 + \frac{1}{\eta}\|\dot{x}(t)\|^2.$$

The conclusion follows in analogy to the proof of (iv) in Theorem 5.11 by using also (5. 75).

(v) Let x^* be the unique element of $\operatorname{zer}(A+B)$. When A is uniformly monotone with corresponding function $\phi_A : [0, +\infty) \to [0, +\infty]$, which is increasing and vanishes only at 0, similarly to the proof of statement (v) in Theorem 5.11 the following inequality can be derived for every $t \in [0, +\infty)$

$$\phi_A\left(\left\|\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|\right) \leq$$
$$\left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t), -B(x(t)) + Bx^* \right\rangle + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle.$$

This yields $\lim_{t\to+\infty} \phi_A\left(\left\|\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|\right) = 0$ and from here the conclusion is immediate.

The case when B is uniformly monotone is to be addressed in analogy to corresponding part of the proof of Theorem 5.11 (v). $\square$

Remark 5.17 In the light of the arguments provided in Remark 5.13, one can see that the statements in Theorem 5.14 remain valid also for $\eta = 2\beta$.

When particularizing this setting to the solving of the optimization problem

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

where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ is a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$, via the second order dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\left[x(t) - \operatorname{prox}_{\eta f}\left(x(t) - \eta\nabla g(x(t))\right)\right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{5. 85}$$

Corollary 5.14 gives rise to the following result.

Corollary 5.8 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ by a proper, convex and lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$ such that* $\operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\} \neq \emptyset$. *Let $\eta \in (0, 2\beta]$ and set $\delta := (4\beta - \eta)/(2\beta)$. Let $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ be functions fulfilling* (A5'), *$u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 85). *Then the following statements are true:*

(i) $x(\cdot)$ *is bounded and* $\dot{x}, \ddot{x}, \left(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\right)x \in L^2([0, +\infty); \mathcal{H})$;

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty}\left(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\right)(x(t)) = 0$;

(iii) $x(t)$ *converges weakly to a minimizer of $f + g$ as $t \to +\infty$;*

(iv) *if x^* is a minimizer of $f+g$, then $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0,+\infty);\mathcal{H})$, $\lim_{t\to+\infty} \nabla g(x(t)) = \nabla g(x^*)$ and ∇g is constant on $\operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\}$;*

(v) *if f or g is uniformly convex, then $x(t)$ converges strongly to the unique minimizer of $f+g$ as $t\to+\infty$.*

As it was also the case in the previous section, we can weaken the choice of the step size in Corollary 5.8 through the following assumption

(A6') $\lambda,\gamma:[0,+\infty)\to(0,+\infty)$ are locally absolutely continuous and there exists $\theta>0$ such that for almost every $t\in[0,+\infty)$ we have

$$\dot{\gamma}(t) \le 0 \le \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \ge \eta\theta + \frac{\eta}{\beta} + 1. \tag{5. 86}$$

Corollary 5.9 *Let $f:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ by a proper, convex and lower semicontinuous function and $g:\mathcal{H}\to\mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta>0$ such that $\operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\}\neq\emptyset$. Let be $\eta>0$, $\lambda,\gamma:[0,+\infty)\to(0,+\infty)$ be functions fulfilling (A6'), $u_0,v_0\in\mathcal{H}$ and $x:[0,+\infty)\to\mathcal{H}$ be the unique strong global solution of (5. 85). Then the following statements are true:*

(i) *$x(\cdot)$ is bounded and $\dot{x},\ddot{x},\big(\operatorname{Id}-\operatorname{prox}_{\eta f}\circ(\operatorname{Id}-\eta\nabla g)\big)x\in L^2([0,+\infty);\mathcal{H})$;*

(ii) *$\lim_{t\to+\infty}\dot{x}(t) = \lim_{t\to+\infty}\ddot{x}(t) = \lim_{t\to+\infty}\big(\operatorname{Id}-\operatorname{prox}_{\eta f}\circ(\operatorname{Id}-\eta\nabla g)\big)(x(t)) = 0$;*

(iii) *$x(t)$ converges weakly to a minimizer of $f+g$ as $t\to+\infty$;*

(iv) *if x^* is a minimizer of $f+g$, then $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0,+\infty);\mathcal{H})$, $\lim_{t\to+\infty} \nabla g(x(t)) = \nabla g(x^*)$ and ∇g is constant on $\operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\}$;*

(v) *if f or g is uniformly convex, then $x(t)$ converges strongly to the unique minimizer of $f+g$ as $t\to+\infty$.*

Proof. The proof follows in the lines of the one given for Corollary 5.5 and relies on the following key inequality, which holds for every $t\in[0,+\infty)$,

$$\begin{aligned}
&\beta\lambda(t)\|\nabla g(x(t)) - \nabla g(x^*)\|^2 + \frac{d^2}{dt^2}\left(\frac{1}{\eta}h+q\right) + \frac{d}{dt}\left(\gamma(t)\left(\frac{1}{\eta}h+q\right)\right)\\
&+\left(\frac{\gamma^2(t)}{\eta\lambda(t)} + \frac{-\dot{\gamma}(t)\lambda(t)+\gamma(t)\dot{\lambda}(t)}{\eta\lambda^2(t)} - \frac{1}{\beta} - \frac{1}{\eta}\right)\|\dot{x}(t)\|^2\\
&+\frac{1}{\eta}\frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right) + \frac{1}{\eta\lambda(t)}\|\ddot{x}(t)\|^2\\
&\le 0,
\end{aligned}$$

where x^* denotes a minimizer of $f+g$. This relation gives rise via (A6') to

$$\begin{aligned}
&\beta\lambda(t)\|\nabla g(x(t)) - \nabla g(x^*)\|^2 + \frac{d}{dt^2}\left(\frac{1}{\eta}h+q\right) + \frac{d}{dt}\left(\gamma(t)\left(\frac{1}{\eta}h+q\right)\right)\\
&+\ \frac{1}{\eta}\frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right) + \theta\|\dot{x}(t)\|^2\ + \frac{1}{\eta\lambda(t)}\|\ddot{x}(t)\|^2\\
&\le 0,
\end{aligned}$$

which can be seen as the counterpart to relation (5. 69). $\square$

Finally, we address the convergence rate of a convex and (Fréchet) differentiable function with Lipschitz continuous gradient $g : \mathcal{H} \to \mathbb{R}$ along the ergodic trajectory generated by

$$\left\{ \begin{array}{l} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\nabla g(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0 \end{array} \right. \tag{5. 87}$$

to its global minimum value, when making the following assumption

(A7') $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous, $\gamma : [0, +\infty) \to (0, +\infty)$ is twice differentiable and there exists $\zeta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$0 < \zeta \leq \gamma(t)\lambda(t) - \dot{\lambda}(t), \ \dot{\gamma}(t) \leq 0 \text{ and } 2\dot{\gamma}(t)\gamma(t) - \ddot{\gamma}(t) \leq 0. \tag{5. 88}$$

Theorem 5.15 *Let $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$ such that* $\operatorname{argmin}_{x \in \mathcal{H}} g(x) \neq \emptyset$*. Let $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ be functions fulfilling* (A7') *$u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of* (5. 87).

Then for every minimizer x^* of g and every $T > 0$ it holds

$$\begin{aligned} 0 &\leq g\left(\frac{1}{T}\int_0^T x(t)dt\right) - g(x^*) \\ &\leq \frac{1}{2\zeta T}\left[\|v_0 + \gamma(0)(u_0 - x^*)\|^2 + \left(\frac{\lambda(0)}{\beta} - \dot{\gamma}(0)\right)\|u_0 - x^*\|^2\right]. \end{aligned}$$

Proof. Let $x^* \in \operatorname{argmin}_{x \in \mathcal{H}} g(x)$ and $T > 0$. By using (5. 87), the convexity of g and (A7') we get for almost every $t \in [0, +\infty)$

$$\begin{aligned} &\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t) + \gamma(t)(x(t) - x^*)\|^2 + \lambda(t)g(x(t)) - \frac{\dot{\gamma}(t)}{2}\|x(t) - x^*\|^2\right) \\ &= \langle \ddot{x}(t) + \dot{\gamma}(t)(x(t) - x^*) + \gamma(t)\dot{x}(t), \dot{x}(t) + \gamma(t)(x(t) - x^*)\rangle \\ &\quad - \frac{\ddot{\gamma}(t)}{2}\|x(t) - x^*\|^2 - \dot{\gamma}(t)\langle \dot{x}(t), x(t) - x^*\rangle + \dot{\lambda}(t)g(x(t)) + \lambda(t)\langle \dot{x}(t), \nabla g(x(t))\rangle \\ &= -\gamma(t)\lambda(t)\langle \nabla g(x(t)), x(t) - x^*\rangle + \dot{\lambda}(t)g(x(t)) + \left(\dot{\gamma}(t)\gamma(t) - \frac{\ddot{\gamma}(t)}{2}\right)\|x(t) - x^*\|^2 \\ &\leq -\gamma(t)\lambda(t)\langle \nabla g(x(t)), x(t) - x^*\rangle + \dot{\lambda}(t)g(x(t)) \\ &\leq (\dot{\lambda}(t) - \gamma(t)\lambda(t))(g(x(t)) - g(x^*)) + \dot{\lambda}(t)g(x^*) \\ &\leq -\zeta(g(x(t)) - g(x^*)) + \dot{\lambda}(t)g(x^*). \end{aligned}$$

We obtain after integration

$$\begin{aligned} &\frac{1}{2}\|\dot{x}(T) + \gamma(T)(x(T) - x^*)\|^2 + \lambda(T)g(x(T)) - \frac{\dot{\gamma}(T)}{2}\|x(T) - x^*\|^2 \\ &- \frac{1}{2}\|\dot{x}(0) + \gamma(0)(x(0) - x^*)\|^2 - \lambda(0)g(x(0)) + \frac{\dot{\gamma}(0)}{2}\|x(0) - x^*\|^2 \\ &+ \zeta\int_0^T (g(x(t)) - g(x^*))dt \\ &\leq (\lambda(T) - \lambda(0))g(x^*). \end{aligned}$$

The conclusion follows from here as in the proof of Theorem 5.12. □

Remark 5.18 A similar comment as in Remark 5.16 can be made also in this context. For $a, a', \rho, \rho' \geq 0$ and $b, b' > 0$ fulfilling the inequalities ${b'}^2 b > 1/\beta$ and $0 \leq \rho \leq b'$ one can prove that the functions

$$\lambda(t) = \frac{1}{ae^{-\rho t} + b} \text{ and } \gamma(t) = a'e^{-\rho' t} + b',$$

verify assumption (A2') in Theorem 5.13 (with $0 < \theta \leq b'^2 b\beta - 1$) and assumption (A7') in Theorem 5.15 (with $0 < \zeta \leq bb'/(a+b)^2$). Hence, for this choice of the relaxation and damping function, one has convergence of the objective function g along the ergodic trajectory to its global minimum value as well as (weak) convergence of the trajectory to a minimizer of g.

5.2.4 Converges rates for strongly monotone inclusions

The starting point of the investigations we go through in this subsection is again the monotone inclusion problem (5. 22), however, this time approached via the second order dynamical system (5. 80).

The following result will be useful when deriving the convergence rates.

Lemma 5.4 *Let $h, \gamma, b_1, b_2, b_3, u : [0, +\infty) \to \mathbb{R}$ be given functions such that h, γ, b_2, u are locally absolutely continuous and $\dot{h}$ is locally absolutely continuous, too. Assume that*

$$h(t), b_2(t), u(t) \geq 0 \ \forall t \in [0, +\infty)$$

and that there exists $\underline{\gamma} > 1$ such that

$$\gamma(t) \geq \underline{\gamma} > 1 \ \forall t \in [0, +\infty).$$

Further, assume that for almost every $t \in [0, +\infty)$ one has

$$\gamma(t) + \dot{\gamma}(t) \leq b_1(t) + 1, \tag{5. 89}$$

$$b_2(t) + \dot{b}_2(t) \leq b_3(t) \tag{5. 90}$$

and

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + b_1(t)h(t) + b_2(t)\dot{u}(t) + b_3(t)u(t) \leq 0. \tag{5. 91}$$

Then there exists $M > 0$ such that the following statements hold:

(i) if $1 < \underline{\gamma} < 2$, then for almost every $t \in [0, +\infty)$

$$0 \leq h(t) \leq \left(h(0) + \frac{M}{2 - \underline{\gamma}} \right) \exp(-(\underline{\gamma} - 1)t);$$

(ii) if $2 < \underline{\gamma}$, then for almost every $t \in [0, +\infty)$

$$\begin{aligned} 0 \leq h(t) &\leq h(0)\exp(-(\underline{\gamma} - 1)t) + \frac{M}{\underline{\gamma} - 2}\exp(-t) \\ &\leq \left(h(0) + \frac{M}{\underline{\gamma} - 2} \right) \exp(-t); \end{aligned}$$

(iii) if $\underline{\gamma} = 2$, then for almost every $t \in [0, +\infty)$

$$0 \leq h(t) \leq (h(0) + Mt)\exp(-t).$$

Proof. We multiply the inequality (5. 91) with $\exp(t)$ and use the identities

$$\begin{aligned} \exp(t)\ddot{h}(t) &= \frac{d}{dt}\big(\exp(t)\dot{h}(t)\big) - \exp(t)\dot{h}(t) \\ \exp(t)\dot{u}(t) &= \frac{d}{dt}(\exp(t)u(t)) - \exp(t)u(t) \\ \exp(t)\dot{h}(t) &= \frac{d}{dt}\big(\exp(t)h(t)\big) - \exp(t)h(t) \end{aligned}$$

in order to derive for almost every $t \in [0,+\infty)$ the inequality

$$\begin{aligned}&\frac{d}{dt}\big(\exp(t)\dot{h}(t)\big) + (\gamma(t)-1)\frac{d}{dt}\big(\exp(t)h(t)\big) + \exp(t)h(t)(b_1(t)+1-\gamma(t))\\&+ b_2(t)\frac{d}{dt}\big(\exp(t)u(t)\big) + (b_3(t)-b_2(t))\exp(t)u(t)\\&\leq 0.\end{aligned}$$

By using also

$$\begin{aligned}(\gamma(t)-1)\frac{d}{dt}\big(\exp(t)h(t)\big) &= \frac{d}{dt}\Big((\gamma(t)-1)\exp(t)h(t)\Big) - \dot{\gamma}(t)\exp(t)h(t)\\ b_2(t)\frac{d}{dt}\big(\exp(t)u(t)\big) &= \frac{d}{dt}\big(b_2(t)\exp(t)u(t)\big) - \dot{b}_2(t)\exp(t)u(t)\end{aligned}$$

we obtain for almost every $t \in [0,+\infty)$

$$\begin{aligned}&\frac{d}{dt}\big(\exp(t)\dot{h}(t)\big) + \frac{d}{dt}\Big((\gamma(t)-1)\exp(t)h(t)\Big) + \frac{d}{dt}\big(b_2(t)\exp(t)u(t)\big)+\\&\big(b_1(t)+1-\gamma(t)-\dot{\gamma}(t)\big)\exp(t)h(t) + \big(b_3(t)-b_2(t)-\dot{b}_2(t)\big)\exp(t)u(t)\\&\leq 0.\end{aligned}$$

The hypotheses regarding the parameters involved imply that the function

$$t \to \exp(t)\dot{h}(t) + (\gamma(t)-1)\exp(t)h(t) + b_2(t)\exp(t)u(t)$$

is monotonically decreasing, hence there exists $M > 0$ such that

$$\exp(t)\dot{h}(t) + (\gamma(t)-1)\exp(t)h(t) + b_2(t)\exp(t)u(t) \leq M.$$

Since $u(t), b_2(t) \geq 0$ we get

$$\dot{h}(t) + (\gamma(t)-1)h(t) \leq M\exp(-t),$$

hence

$$\dot{h}(t) + (\underline{\gamma}-1)h(t) \leq M\exp(-t)$$

for every $t \in [0,+\infty)$. This implies that

$$\frac{d}{dt}\big(\exp((\underline{\gamma}-1)t)h(t)\big) \leq M\exp((\underline{\gamma}-2)t),$$

for every $t \in [0,+\infty)$, from which the conclusion follows easily by integration. $\square$

We come now to the first main result of this section.

Theorem 5.16 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $B : \mathcal{H} \to \mathcal{H}$ a monotone and $\frac{1}{\beta}$-Lipschitz continuous operator for $\beta > 0$ such that $A + B$ is ρ-strongly monotone for $\rho > 0$ and x^* be the unique point in $\operatorname{zer}(A+B)$. Chose $\alpha, \delta \in (0,1)$ and $\eta > 0$ such that $\delta\beta\rho < 1$ and $\frac{1}{\eta} = \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta} - \rho > 0$.*

Let $\lambda : [0,+\infty) \to [0,+\infty)$ be a locally absolutely continuous function fulfilling for every $t \in [0,+\infty)$

$$\begin{aligned}\theta(t) &:= \lambda(t)\frac{\delta}{1-\delta}\frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}}\\&\leq \lambda(t)\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} + \lambda^2(t)\left(\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}\right)^2\end{aligned}$$

and such that there exists a real number $\underline{\lambda}$ *with the property that*

$$0 < \underline{\lambda} \leq \inf_{t \geq 0} \lambda(t)$$

and

$$2 < \theta := \underline{\lambda} \frac{\delta}{1-\delta} \frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}}.$$

Further, let $\gamma : [0, +\infty) \to [0, +\infty)$ *be a locally absolutely continuous function fulfilling*

$$\frac{1 + \sqrt{1 + 4\theta(t)}}{2} \leq \gamma(t) \leq 1 + \lambda(t) \frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} \text{ for every } t \in [0, +\infty) \tag{5. 92}$$

and

$$\dot{\gamma}(t) \leq 0 \text{ and } \frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\right) \leq 0 \text{ for almost every } t \in [0, +\infty). \tag{5. 93}$$

Let $u_0, v_0 \in \mathcal{H}$ *and* $x : [0, +\infty) \to \mathcal{H}$ *be the unique strong global solution of the dynamical system* (5. 80).

Then $\gamma(t) \geq \underline{\gamma} := \frac{1+\sqrt{1+4\theta}}{2} > 2$ *for every* $t \in [0, +\infty)$ *and there exists* $M > 0$ *such that for every* $t \in [0, +\infty)$

$$\begin{aligned} 0 \leq \|x(t) - x^*\|^2 &\leq \|u_0 - x^*\|^2 \exp(-(\underline{\gamma} - 1)t) + \frac{M}{\underline{\gamma} - 2} \exp(-t) \\ &\leq \left(\|u_0 - x^*\|^2 + \frac{M}{\underline{\gamma} - 2}\right) \exp(-t). \end{aligned}$$

Proof. From the definition of the resolvent we have for almost every $t \in [0, +\infty)$

$$\begin{aligned} &B\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right) - B(x(t)) - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \\ &\in (A+B)\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right). \end{aligned} \tag{5. 94}$$

We combine this with $0 \in (A+B)x^*$, the strong monotonicity of $A+B$, the Lipschitz continuity of B and, by also using the Cauchy-Schwartz inequality, we get for almost every $t \in [0,+\infty)$

$$\begin{aligned}
&\frac{\rho}{\lambda^2(t)}\left\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\right\|^2+\frac{2\rho}{\lambda(t)}\left\langle x(t)-x^*, \ddot{x}(t)+\gamma(t)\dot{x}(t)\right\rangle+\rho\|x(t)-x^*\|^2\\
&=\rho\left\|\frac{1}{\lambda(t)}\ddot{x}(t)+\frac{\gamma(t)}{\lambda(t)}\dot{x}(t)+x(t)-x^*\right\|^2\\
&\leq\left\langle\frac{1}{\lambda(t)}\ddot{x}(t)+\frac{\gamma(t)}{\lambda(t)}\dot{x}(t)+x(t)-x^*, B\left(\frac{1}{\lambda(t)}\ddot{x}(t)+\frac{\gamma(t)}{\lambda(t)}\dot{x}(t)+x(t)\right)-B(x(t))\right\rangle\\
&\quad-\left\langle\frac{1}{\lambda(t)}\ddot{x}(t)+\frac{\gamma(t)}{\lambda(t)}\dot{x}(t)+x(t)-x^*, \frac{1}{\eta\lambda(t)}\ddot{x}(t)+\frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t)\right\rangle\\
&=\frac{1}{\lambda(t)}\left\langle\ddot{x}(t)+\gamma(t)\dot{x}(t), B\left(\frac{1}{\lambda(t)}\ddot{x}(t)+\frac{\gamma(t)}{\lambda(t)}\dot{x}(t)+x(t)\right)-B(x(t))\right\rangle\\
&\quad+\left\langle x(t)-x^*, B\left(\frac{1}{\lambda(t)}\ddot{x}(t)+\frac{\gamma(t)}{\lambda(t)}\dot{x}(t)+x(t)\right)-B(x(t))\right\rangle\\
&\quad-\frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2-\frac{1}{\eta\lambda(t)}\left\langle x(t)-x^*, \ddot{x}(t)+\gamma(t)\dot{x}(t)\right\rangle\\
&\leq\frac{1}{\beta\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2-\frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2\\
&\quad+\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2+\rho\alpha\|x(t)-x^*\|^2\\
&\quad-\frac{1}{\eta\lambda(t)}\left\langle x(t)-x^*, \ddot{x}(t)+\gamma(t)\dot{x}(t)\right\rangle.
\end{aligned}$$

Using again the notation $h(t)=\frac{1}{2}\|x(t)-x^*\|^2$, we have for almost every $t \in [0,+\infty)$

$$\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2=\|\ddot{x}(t)\|^2+\gamma^2(t)\|\dot{x}(t)\|^2+\gamma(t)\frac{d}{dt}(\|\dot{x}(t)\|^2) \tag{5. 95}$$

and

$$\langle x(t)-x^*, \ddot{x}(t)+\gamma(t)\dot{x}(t)\rangle=\ddot{h}(t)+\gamma(t)\dot{h}(t)-\|\dot{x}(t)\|^2.$$

Therefore, we obtain for almost every $t \in [0,+\infty)$

$$\begin{aligned}
&\left(\frac{\rho}{\lambda^2(t)}+\frac{1}{\eta\lambda^2(t)}-\frac{1}{\beta\lambda^2(t)}-\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}\right)\|\ddot{x}(t)\|^2\\
&+\left[\gamma^2(t)\left(\frac{\rho}{\lambda^2(t)}+\frac{1}{\eta\lambda^2(t)}-\frac{1}{\beta\lambda^2(t)}-\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}\right)-\frac{2\rho}{\lambda(t)}-\frac{1}{\eta\lambda(t)}\right]\|\dot{x}(t)\|^2\\
&+\gamma(t)\left(\frac{\rho}{\lambda^2(t)}+\frac{1}{\eta\lambda^2(t)}-\frac{1}{\beta\lambda^2(t)}-\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}\right)\frac{d}{dt}\left(\|\dot{x}(t)\|^2\right)\\
&+\left(\frac{2\rho}{\lambda(t)}+\frac{1}{\eta\lambda(t)}\right)\ddot{h}(t)+\gamma(t)\left(\frac{2\rho}{\lambda(t)}+\frac{1}{\eta\lambda(t)}\right)\dot{h}(t)+2\rho(1-\alpha)h(t)\\
&\leq 0.
\end{aligned}$$

The hypotheses imply that

$$\frac{\rho}{\lambda^2(t)}+\frac{1}{\eta\lambda^2(t)}-\frac{1}{\beta\lambda^2(t)}-\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}=\frac{1}{\lambda^2(t)}\left(\rho+\frac{1}{\eta}-\frac{1}{\beta}-\frac{1}{4\rho\beta^2\alpha}\right)>0,$$

hence the first term in the left hand side of the above inequality can be neglected and we obtain for almost every $t \in [0,+\infty)$ that

$$\ddot{h}(t)+\gamma(t)\dot{h}(t)+b_1(t)h(t)+b_2(t)\frac{d}{dt}(\|\dot{x}(t)\|^2)+b_3(t)\|\dot{x}(t)\|^2\leq 0, \tag{5. 96}$$

where

$$b_1(t) := \lambda(t)\frac{2\rho(1-\alpha)}{2\rho+\frac{1}{\eta}} > 0,$$

$$b_2(t) := \gamma(t)\frac{\frac{\rho}{\lambda^2(t)}+\frac{1}{\eta\lambda^2(t)}-\frac{1}{\beta\lambda^2(t)}-\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}}{\frac{2\rho}{\lambda(t)}+\frac{1}{\eta\lambda(t)}} = \frac{\gamma(t)}{\lambda(t)}\frac{\rho+\frac{1}{\eta}-\frac{1}{\beta}-\frac{1}{4\rho\beta^2\alpha}}{2\rho+\frac{1}{\eta}} > 0$$

and

$$b_3(t) := \frac{\gamma^2(t)\left(\frac{\rho}{\lambda^2(t)}+\frac{1}{\eta\lambda^2(t)}-\frac{1}{\beta\lambda^2(t)}-\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}\right)-\frac{2\rho}{\lambda(t)}-\frac{1}{\eta\lambda(t)}}{\frac{2\rho}{\lambda(t)}+\frac{1}{\eta\lambda(t)}}.$$

This shows that (5. 91) in Lemma 5.4 for $u := \|\dot{x}(\cdot)\|^2$ is fulfilled. In order to apply Lemma 5.4, we have only to prove that (5. 89) and (5. 90) are satisfied, as every other assumption in this statement is obviously guaranteed.

A simple calculation shows that

$$b_3(t) \geq b_2(t) \Longleftrightarrow \gamma^2(t)-\gamma(t) \geq \frac{\frac{2\rho}{\lambda(t)}+\frac{1}{\eta\lambda(t)}}{\frac{\rho}{\lambda^2(t)}+\frac{1}{\eta\lambda^2(t)}-\frac{1}{\beta\lambda^2(t)}-\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}} = \theta(t), \tag{5. 97}$$

which is true according to (5. 92), thus $b_3(t) \geq b_2(t)$ for every $t \in [0,+\infty)$. On the other hand (see (5. 93)),

$$\dot{b}_2(t) \leq 0$$

for almost every $t \in [0,+\infty)$, from which (5. 90) follows.

Further, again by using (5. 92), observe that

$$1+b_1(t) = 1+\lambda(t)\frac{2\rho(1-\alpha)}{\rho+\left(\frac{1}{\beta}+\frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} \geq \gamma(t)$$

for every $t \in [0,+\infty)$, which, combined with

$$\dot{\gamma}(t) \leq 0$$

for almost every $t \in [0,+\infty)$, shows that (5. 89) is also fulfilled.

The conclusion follows from Lemma 5.4(ii), by noticing that $\underline{\gamma} > 2$, as $\theta > 2$. □

Remark 5.19 One can notice that when $\dot{\gamma}(t) \leq 0$ for almost every $t \in [0,+\infty)$, the second assumption in (5. 93) is fulfilled provided that $\dot{\lambda}(t) \geq 0$ for almost every $t \in [0,+\infty)$.

Further, we would like to point out that one can obviously chose $\lambda(t) = \underline{\lambda}$ and $\gamma(t) = \gamma$ for every $t \in [0,+\infty)$, where

$$\begin{aligned} 2 < \theta &:= \underline{\lambda}\frac{\delta}{1-\delta}\frac{\rho+\left(\frac{1}{\beta}+\frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta}+\frac{1}{4\rho\beta^2\alpha}} \\ &\leq \underline{\lambda}\frac{2\rho(1-\alpha)}{\rho+\left(\frac{1}{\beta}+\frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} + \underline{\lambda}^2\left(\frac{2\rho(1-\alpha)}{\rho+\left(\frac{1}{\beta}+\frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}\right)^2 \end{aligned}$$

and

$$\frac{1+\sqrt{1+4\theta}}{2} \leq \gamma \leq 1+\underline{\lambda}\frac{2\rho(1-\alpha)}{\rho+\left(\frac{1}{\beta}+\frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}.$$

When considering the convex optimization problem (5. 26), the second order dynamical system (5. 80) written for $A = \partial f$ and $B = \nabla g$ becomes

$$\left\{ \begin{array}{l} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\left[x(t) - \operatorname{prox}_{\eta f}\Big(x(t) - \eta\nabla g(x(t))\Big)\right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{array} \right. \tag{5. 98}$$

Theorem 5.16 gives rise to the following result.

Theorem 5.17 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function, $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\beta > 0$ such that $f + g$ is ρ-strongly convex for $\rho > 0$ and x^* be the unique minimizer of $f + g$ over $\mathcal{H}$. Chose $\alpha, \delta \in (0,1)$ and $\eta > 0$ such that $\delta\beta\rho < 1$ and $\frac{1}{\eta} = \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta} - \rho > 0$.*

Let $\lambda : [0,+\infty) \to [0,+\infty)$ be a locally absolutely continuous function fulfilling for every $t \in [0,+\infty)$

$$\begin{aligned} \theta(t) &:= \lambda(t)\frac{\delta}{1-\delta}\frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}} \\ &\leq \lambda(t)\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} + \lambda^2(t)\left(\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}\right)^2 \end{aligned}$$

and such that there exists a real number $\underline{\lambda}$ with the property that

$$0 < \underline{\lambda} \leq \inf_{t \geq 0} \lambda(t)$$

and

$$2 < \theta := \underline{\lambda}\frac{\delta}{1-\delta}\frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}}.$$

Further, let $\gamma : [0,+\infty) \to [0,+\infty)$ be a locally absolutely continuous function fulfilling (5. 92) *and* (5. 93)*, $u_0, v_0 \in \mathcal{H}$ and $x : [0,+\infty) \to \mathcal{H}$ be the unique strong global solution of the dynamical system* (5. 98)*.*

Then $\gamma(t) \geq \underline{\gamma} := \frac{1+\sqrt{1+4\theta}}{2} > 2$ for every $t \in [0,+\infty)$ and there exists $M > 0$ such that for every $t \in [0,+\infty)$

$$\begin{aligned} 0 \leq \|x(t) - x^*\|^2 &\leq \|u_0 - x^*\|^2 \exp(-(\underline{\gamma} - 1)t) + \frac{M}{\underline{\gamma} - 2}\exp(-t) \\ &\leq \left(\|u_0 - x^*\|^2 + \frac{M}{\underline{\gamma} - 2}\right)\exp(-t). \end{aligned}$$

Finally, we approach the convex minimization problem (5. 28) via the second order dynamical system

$$\left\{ \begin{array}{l} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\nabla g(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0 \end{array} \right. \tag{5. 99}$$

and provide an exponential rate of convergence of g to its minimum value along the generated trajectories. The following result can be seen as the continuous counterpart of [89, Theorem 4], where recently a linear rate of convergence for the values of g on a sequence iteratively generated by an inertial-type algorithm has been obtained.

Theorem 5.18 *Let $g : \mathcal{H} \to \mathbb{R}$ be a ρ-strongly convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\rho > 0$ and $\beta > 0$ and x^* be the unique minimizer of g over $\mathcal{H}$.*

Let $\alpha : [0, +\infty) \to \mathbb{R}$ be a Lebesgue measurable function such that there exists $\underline{\alpha} > 1$ with

$$\inf_{t \geq 0} \alpha(t) \geq \max\left\{\underline{\alpha}, \frac{2}{\beta^2\rho^2} - 1\right\} \tag{5. 100}$$

and $\lambda : [0, +\infty) \to [0, +\infty)$ be a locally absolutely continuous function fulfilling for every $t \in [0, +\infty)$

$$\frac{\alpha(t)}{\beta\rho^2} \leq \lambda(t) \leq \frac{\beta}{2}\big(\alpha(t) + \alpha^2(t)\big). \tag{5. 101}$$

Further, let $\gamma : [0, +\infty) \to [0, +\infty)$ be a locally absolutely continuous function fulfilling

$$\frac{1 + \sqrt{1 + 8\frac{\lambda(t)}{\beta}}}{2} \leq \gamma(t) \leq 1 + \alpha(t) \text{ for every } t \in [0, +\infty) \tag{5. 102}$$

and (5. 93)*.*

Let $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of the dynamical system (5. 99)*.*

Then $\gamma(t) \geq \underline{\gamma} := \frac{1+\sqrt{1+8\frac{\underline{\alpha}}{\beta^2\rho^2}}}{2} > 2$ and there exists $M > 0$ such that for every $t \in [0, +\infty)$

$$\begin{aligned}
0 \leq \frac{\rho}{2}\|x(t) - x^*\|^2 &\leq g(x(t)) - g(x^*) \\
&\leq (g(u_0) - g(x^*))\exp(-(\underline{\gamma} - 1)t) + \frac{M}{\underline{\gamma} - 2}\exp(-t) \\
&\leq \left(g(u_0) - g(x^*) + \frac{M}{\underline{\gamma} - 2}\right)\exp(-t) \\
&\leq \left(\frac{1}{2\beta}\|u_0 - x^*\|^2 + \frac{M}{\underline{\gamma} - 2}\right)\exp(-t).
\end{aligned}$$

Proof. One has for almost every $t \in [0, +\infty)$

$$\frac{d}{dt}g(x(t)) = \langle \dot{x}(t), \nabla g(x(t))\rangle$$

and (see Remark 5.1(b))

$$\frac{d^2}{dt^2}g(x(t)) = \langle \ddot{x}(t), \nabla g(x(t))\rangle + \left\langle \dot{x}(t), \frac{d}{dt}\nabla g(x(t))\right\rangle \leq \langle \ddot{x}(t), \nabla g(x(t))\rangle + \frac{1}{\beta}\|\dot{x}(t)\|^2.$$

Further, by using (5. 30), (5. 31) and the first equation in (5. 99), we derive for almost every $t \in [0, +\infty)$

$$\begin{aligned}
&\frac{d^2}{dt^2}\big(g(x(t)) - g(x^*)\big) + \gamma(t)\frac{d}{dt}\big(g(x(t)) - g(x^*)\big) + \alpha(t)\big(g(x(t)) - g(x^*)\big) \\
&\leq\ -\lambda(t)\|\nabla g(x(t))\|^2 + \frac{\alpha(t)}{2\beta\rho^2}\|\nabla g(x(t))\|^2 + \frac{1}{\beta}\|\dot{x}(t)\|^2 \\
&= -\frac{1}{2\lambda(t)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 - \frac{\lambda(t)}{2}\|\nabla g(x(t))\|^2 \\
&\quad + \frac{\alpha(t)}{2\beta\rho^2}\|\nabla g(x(t))\|^2 + \frac{1}{\beta}\|\dot{x}(t)\|^2.
\end{aligned}$$

Taking into account (5. 95) we obtain for almost every $t \in [0, +\infty)$

$$\begin{aligned}
&\frac{d^2}{dt^2}\big(g(x(t)) - g(x^*)\big) + \gamma(t)\frac{d}{dt}\big(g(x(t)) - g(x^*)\big) + \alpha(t)\big(g(x(t)) - g(x^*)\big)\\
+\ &\frac{\gamma(t)}{2\lambda(t)}\frac{d}{dt}(\|\dot{x}(t)\|^2) + \left(\frac{\gamma^2(t)}{2\lambda(t)} - \frac{1}{\beta}\right)\|\dot{x}(t)\|^2\\
+\ &\frac{1}{2\lambda(t)}\|\ddot{x}(t)\|^2 + \left(\frac{\lambda(t)}{2} - \frac{\alpha(t)}{2\beta\rho^2}\right)\|\nabla g(x(t))\|^2\\
\le\ &0.
\end{aligned}$$

According to the choice of the parameters involved, we have

$$\frac{\lambda(t)}{2} - \frac{\alpha(t)}{2\beta\rho^2} \ge 0,$$

thus, for almost every $t \in [0, +\infty)$,

$$\begin{aligned}
&\frac{d^2}{dt^2}\big(g(x(t)) - g(x^*)\big) + \gamma(t)\frac{d}{dt}\big(g(x(t)) - g(x^*)\big) + \alpha(t)\big(g(x(t)) - g(x^*)\big)\\
+\ &\frac{\gamma(t)}{2\lambda(t)}\frac{d}{dt}(\|\dot{x}(t)\|^2) + \left(\frac{\gamma^2(t)}{2\lambda(t)} - \frac{1}{\beta}\right)\|\dot{x}(t)\|^2\\
\le\ &0.
\end{aligned}$$

This shows that (5. 91) in Lemma 5.4 for $u := \|\dot{x}(\cdot)\|^2$,

$$b_1(t) := \alpha(t),$$

$$b_2(t) := \frac{\gamma(t)}{2\lambda(t)}$$

and

$$b_3(t) := \frac{\gamma^2(t)}{2\lambda(t)} - \frac{1}{\beta}$$

is fulfilled. By combining (5. 102) and the first condition in (5. 93) one obtains (5. 89), while, by combining (5. 102) and the second condition in (5. 93) one obtains (5. 90).

Furthermore, by taking into account the Lipschitz property of ∇g and the strong convexity of g, it yields

$$\rho\beta \le 1.$$

From (5. 101), (5. 100) and $\underline{\alpha} > 1$ we obtain

$$\frac{\lambda(t)}{\beta} \ge \underline{\alpha}\frac{1}{\beta^2\rho^2} > 1 \text{ for every } t \in [0, +\infty),$$

which combined with (5. 102) leads to $\underline{\gamma} > 2$.

The conclusion follows from Lemma 5.4(ii), the strong convexity of g and (5. 30).
□

Remark 5.20 In Theorem 5.18 one can obviously chose $\alpha(t) = \alpha$, where $\alpha = \frac{2}{\beta^2\rho^2} - 1$, if $\beta\rho < 1$, or $\alpha = 1 + \varepsilon$, with $\varepsilon > 0$, otherwise, $\lambda(t) = \lambda$ and $\gamma(t) = \gamma$ for every $t \in [0, +\infty)$, where

$$\frac{\alpha}{\beta\rho^2} \le \lambda \le \frac{\beta}{2}\big(\alpha + \alpha^2\big)$$

and

$$\frac{1 + \sqrt{1 + 8\frac{\lambda}{\beta}}}{2} \le \gamma \le 1 + \alpha.$$

Bibliography

[1] B. Abbas, H. Attouch, *Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator*, Optimization 64(10), 2223–2252, 2015

[2] B. Abbas, H. Attouch, B.F. Svaiter, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, Journal of Optimization Theory and its Applications 161(2), 331–360, 2014

[3] 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

[4] 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

[5] 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

[6] F. Alvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, Journal de Mathématiques Pures et Appliquées (9) 81(8), 747–779, 2002

[7] A.S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, (Russian) Differentsial'nye Uravneniya 30(9), 1475–1486, 1994; translation in Differential Equations 30(9), 1365–1375, 1994

[8] H. Attouch, F. Alvarez, *The heavy ball with friction dynamical system for convex constrained minimization problems*, in: Optimization (Namur, 1998), 25–35, in: Lecture Notes in Economics and Mathematical Systems 481, Springer, Berlin, 2000

[9] H. Attouch, J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Mathematical Programming 116(1-2) Series B, 5–16, 2009

[10] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality*, Mathematics of Operations Research 35(2), 438–457, 2010

[11] H. Attouch, J. Bolte, B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Mathematical Programming 137(1-2) Series A, 91–129, 2013

[12] H. Attouch, L.M. Briceño-Arias, P.L. Combettes, *A parallel splitting method for coupled monotone inclusions*, SIAM Journal on Control and Optimization 48(5), 3246–3270, 2010

[13] H. Attouch, Z. Chbani, *Fast inertial dynamics and FISTA algorithms in convex optimization. Perturbation aspects*, arXiv:1507.01367, 2015

[14] H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, Journal of Differential Equations 248(6), 1315-1344, 2010

[15] H. Attouch, M.-O. Czarnecki, J. Peypouquet, *Prox-penalization and splitting methods for constrained variational problems*, SIAM Journal on Optimization 21(1), 149-173, 2011

[16] H. Attouch, M.-O. Czarnecki, J. Peypouquet, *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM Journal on Optimization 21(4), 1251-1274, 2011

[17] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system*, Communications in Contemporary Mathematics 2(1), 1–34, 2000

[18] H. Attouch, M. Marques Alves, B.F. Svaiter, *A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert spaces, with complexity $O(1/n^2)$*, Journal of Convex Analysis 23(1), 139–180, 2016

[19] 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

[20] H. Attouch, B.F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM Journal on Control and Optimization 49(2), 574–598, 2011

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

[22] J.-P. Aubin, A. Cellina, *Differential inclusions. Set-valued maps and viability theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] 264, Springer-Verlag, Berlin, 1984

[23] J.B. Baillon, H. Brézis, *Une remarque sur le comportement asymptotique des semigroupes non linéaires*, Houston Journal of Mathematics 2(1), 5–7, 1976

[24] S. Banert, R.I. Boţ, *Backward penalty schemes for monotone inclusion problems*, Journal of Optimization Theory and Applications 166(3), 930–948, 2015

[25] S. Banert, R.I. Boţ, *A forward-backward-forward differential equation and its asymptotic properties*, arXiv:1503.07728, 2015

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

[27] H.H. Bauschke, D.A. McLaren, H.S. Sendov, *Fitzpatrick functions: inequalities, examples and remarks on a problem by S. Fitzpatrick*, Journal of Convex Analysis 13(3-4), 499–523, 2006

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

[29] S.R. Becker, P.L. Combettes, *An algorithm for splitting parallel sums of linearly composed monotone operators with applications to signal recovery*, Journal of Nonlinear and Convex Analysis 15(1), 137–159, 2014

[30] D.P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Cambridge, MA, 1999

[31] J. Bolte, *Continuous gradient projection method in Hilbert spaces*, Journal of Optimization Theory and its Applications 119(2), 235–259, 2003

[32] J. Bolte, A. Daniilidis, A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization 17(4), 1205–1223, 2006

[33] J. Bolte, A. Daniilidis, A. Lewis, M. Shota, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization 18(2), 556–572, 2007

[34] J. Bolte, A. Daniilidis, O. Ley, L. Mazet, *Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society 362(6), 3319–3363, 2010

[35] J. Bolte, S. Sabach, M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming Series A (146)(1–2), 459–494, 2014

[36] J.M. Borwein, *Maximal monotonicity via convex analysis*, Journal of Convex Analysis 13(3-4), 561–586, 2006

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

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

[39] R.I. Boţ, E.R. Csetnek, *An application of the bivariate inf-convolution formula to enlargements of monotone operators*, Set-Valued Analysis 16(7-8), 983–997, 2008

[40] 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

[41] R.I. Boţ, E.R. Csetnek, *Forward-backward and Tseng's type penalty schemes for monotone inclusion problems*, Set-Valued and Variational Analysis 22, 313–331, 2014

[42] R.I. Boţ, E.R. Csetnek, *A Tseng's type penalty scheme for solving inclusion problems involving linearly composed and parallel-sum type monotone operators*, Vietnam Journal of Mathematics 42(4), 451–465, 2014

[43] R.I. Boţ, E.R. Csetnek, *On the convergence rate of a forward-backward type primal-dual splitting algorithm for convex optimization problems*, Optimization 64(1), 5–23, 2015

[44] R.I. Boţ, E.R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numerical Algorithms 71, 519–540, 2016

[45] R.I. Boţ, E.R. Csetnek, *An inertial alternating direction method of multipliers*, Minimax Theory and its Applications 1(1), 29–49, 2016

[46] R.I. Boţ, E.R. Csetnek, *A hybrid proximal-extragradient algorithm with inertial effects*, Numerical Functional Analysis and Optimization 36(8), 951–963, 2015

[47] R.I. Boţ, E.R. Csetnek, *An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems*, Journal of Optimization Theory and Applications, DOI 10.1007/s10957-015-0730-z

[48] R.I. Boţ, E.R. Csetnek, *A dynamical system associated with the fixed points set of a nonexpansive operator*, Journal of Dynamics and Differential Equations, DOI: 10.1007/s10884-015-9438-x, 2015

[49] R.I. Boţ, E.R. Csetnek, *Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems*, Journal of Mathematical Analysis and Applications 435(2), 1688–1700, 2016

[50] R.I. Boţ, E.R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, arXiv:1503.04652, 2015

[51] R.I. Boţ, E.R. Csetnek, *Convergence rates for forward-backward dynamical systems associated with strongly monotone inclusions*, arXiv:1504.01863, 2015

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

[53] 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 150(2), 251–279, 2015

[54] R.I. Boţ, E.R. Csetnek, C. Hendrich, *Recent developments on primal-dual splitting methods with applications to convex minimization*, in: P.M. Pardalos, T.M. Rassias (Eds.), "Mathematics Without Boundaries: Surveys in Interdisciplinary Research", Springer-Verlag, New York, 2014

[55] R.I. Boţ, E.R. Csetnek, C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Applied Mathematics and Computation 256, 472–487, 2015

[56] R.I. Boţ, E.R. Csetnek, S. László, *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions*, EURO Journal on Computational Optimization 4, 3–25, 2016

[57] R.I. Boţ, C. Hendrich, *Solving monotone inclusions involving parallel sums of linearly composed maximally monotone operators*, arXiv:1306.3191v2, 2013

[58] 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

[59] 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

[60] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Mathematics Studies No. 5, Notas de Matemática (50), North-Holland/Elsevier, New York, 1973

[61] L.M. Briceño-Arias, *Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions*, Optimization 64(5), 1239-1261, 2015

[62] 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

[63] R.E. Bruck, Jr., *Asymptotic convergence of nonlinear contraction semigroups in Hilbert space*, Journal of Functional Analysis 18, 15–26, 1975

[64] R.S. Burachik, B.F. Svaiter, *Maximal monotone operators, convex functions and a special family of enlargements*, Set-Valued Analysis 10(4), 297–316, 2002

[65] A. Cabot, H. Engler, S. Gadat, *On the long time behavior of second order differential equations with asymptotically small dissipation*, Transactions of the American Mathematical Society 361(11), 5983–6017, 2009

[66] A. Cabot, H. Engler, S. Gadat, *Second-order differential equations with asymptotically small dissipation and piecewise flat potentials*, Proceedings of the Seventh Mississippi StateUAB Conference on Differential Equations and Computational Simulations, 33–38, Electronic Journal of Differential Equations Conference 17, 2009

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

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

[69] 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

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

[71] C. Chen, S. MA, J. Yang, *A general inertial proximal point method for mixed variational inequality problem*, SIAM Journal on Optimization 25(4), 2120–2142, 2015

[72] Y. Chen, G. Lan, Y. Ouyang, *Optimal primal-dual methods for a class of saddle point problems*, SIAM Journal on Optimization 24(4), 1779–1814, 2014

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

[74] E. Chouzenoux, J.-C. Pesquet, A. Repetti, *Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function*, Journal of Optimization Theory and its Applications 162(1), 107–132, 2014

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

[76] 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

[77] P.L. Combettes, I. Yamada, *Compositions and convex combinations of averaged nonexpansive operators*, Journal of Mathematical Analysis and Applications 425(1), 55–70, 2015

[78] P.L. Combettes, V.R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation 4(4), 1168–1200, 2005

[79] 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

[80] E. Corman, X. Yuan, *A Generalized proximal point algorithm and its convergence rate*, SIAM Journal on Optimization, 24(4), 1614–1638, 2014

[81] D. Davis, *Convergence rate analysis of primal-dual splitting schemes*, SIAM Journal on Optimization 25(3), 1912–1943, 2015

[82] D. Davis, W. Yin, *Convergence rate analysis of several splitting schemes*, arXiv: 1406.4834v3, 2015

[83] 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

[84] J. Eckstein, D.P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming 55, 293–318, 1992

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

[86] E. Esser, X. Zhang, T.F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences 3(4), 1015-1046, 2010

[87] S. Fitzpatrick, *Representing monotone operators by convex functions*, in: Workshop/Miniconference on Functional Analysis and Optimization (Canberra, 1988), Proceedings of the Centre for Mathematical Analysis **20**, Australian National University, Canberra, 59–65, 1988

[88] P. Frankel, G. Garrigos, J. Peypouquet, *Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates*, Journal of Optimization Theory and its Applications 165(3), 874–900, 2015

[89] E. Ghadimi, H.R. Feyzmahdavian, M. Johansson, *Global convergence of the Heavy-ball method for convex optimization*, arXiv:1412.7457, 2014

[90] A. Haraux, *Systèmes Dynamiques Dissipatifs et Applications*, Recherches en Mathé- matiques Appliquées 17, Masson, Paris, 1991

[91] R. Hesse, D.R. Luke, S. Sabach, M.K. Tam, *Proximal heterogeneous block input-output method and application to blind ptychographic diffraction imaging*, SIAM Journal on Imaging Sciences 8(1), 426–457, 2015

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

[93] N.V. Krylov, *Some properties of monotone mappings*, (Russian) Litovsk. Mat. Sb. 22(2), 80–87, 1982

[94] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Annales de l'institut Fourier (Grenoble) 48(3), 769–783, 1998

[95] J. Liang, J. Fadili, G. Peyré, *Convergence rates with inexact nonexpansive operators*, arXiv:1404.4837, 2014

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

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

[98] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique Paris, 87–89, 1963

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

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

[101] B. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*, Springer-Verlag, Berlin, 2006

[102] 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

[103] 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

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

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

[106] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, Dordrecht, 2004

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

[108] D. Noll, *Convergence of non-smooth descent methods using the Kurdyka-Lojasiewicz inequality*, Journal of Optimization Theory and Applications 160(2), 553–572, 2014

[109] N. Noun, J. Peypouquet, *Forward-backward penalty scheme for constrained convex minimization without inf-compactness*, Journal of Optimization Theory and Applications, 158(3), 787–795, 2013

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

[111] N. Ogura, I. Yamada, *Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping*, Numerical Functional Analysis and Optimization 23(1-2), 113–137, 2002

[112] Opial, Z., *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bulletin of the American Mathematical Society 73, 591–597, 1967

[113] Passty, G., *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, Journal of Mathematical Analysis and Applications 72, 383–390, 1979

[114] A. Pazy, *Semigroups of nonlinear contractions and their asymptotic behaviour*, in Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. III (Heriot-Watt Univ., Edinburgh), R.J. Knops (ed.), pp. 36–134, Res. Notes in Math., 30, Pitman, Boston, Mass.-London, 1979

[115] J. Peypouquet, *Coupling the gradient method with a general exterior penalization scheme for convex minimization*, Journal of Optimization Theory and Applications 153(1), 123-138, 2012

[116] J.-C. Pesquet, N. Pustelnik, *A parallel inertial proximal optimization method*, Pacific Journal of Optimization 8(2), 273–305, 2012

[117] J. Peypouquet, S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, Journal of Convex Analysis 17(3-4), 1113–1163, 2010

[118] B.T. Polyak, *Introduction to Optimization*, (Translated from the Russian) Translations Series in Mathematics and Engineering, Optimization Software, Inc., Publications Division, New York, 1987

[119] H. Raguet, J. Fadili, G. Peyré, *A generalized forward-backward splitting*, SIAM Journal on Imaging Sciences 6(3), 1199–1226, 2013

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

[121] R.T. Rockafellar, *On the maximality of sums of nonlinear monotone operators*, Transactions of the American Mathematical Society 149, 75–88, 1970

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

[123] R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis*, Fundamental Principles of Mathematical Sciences 317, Springer-Verlag, Berlin, 1998

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

[125] E.D. Sontag, *Mathematical control theory. Deterministic finite-dimensional systems*, Second edition, Texts in Applied Mathematics 6, Springer-Verlag, New York, 1998

[126] W. Su, S. Boyd, E.J. Candes, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*, arXiv:1503.01243, 2015

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

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

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

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

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

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