A DYNAMICAL APPROACH TO AN INERTIAL FORWARD-BACKWARD ALGORITHM FOR CONVEX MINIMIZATION

HÉDY ATTOUCH, JUAN PEYPOUQUET, AND PATRICK REDONT

Abstract. We introduce a new class of forward-backward algorithms for structured convex minimization problems in Hilbert spaces. Our approach relies on the time discretization of a second-order differential system with two potentials and Hessian-driven damping, recently introduced in [11]. This system can be equivalently written as a first-order system in time and space, each of the two constitutive equations involving one (and only one) of the two potentials. Its time dicretization naturally leads to the introduction of forward-backward splitting algorithms with inertial features. Using a Liapunov analysis, we show the convergence of the algorithm under conditions enlarging the classical step size limitation. Then, we specialize our results to gradient-projection algorithms, and give some illustration to sparse signal recovery and feasibility problems.

Key words: Inertial forward-backward algorithm, Dynamical approach, Convex minimization
MSC2010: 90C25, 49M37, 65K05, 47N10

Introduction

Forward-backward algorithms and their natural companions, the gradient-projection algorithms, have proved to be efficient tools for solving structured optimization problems. They provide parallel splitting methods which are particularly interesting for large-scale systems. In this paper, we introduce a new class of forward-backward algorithms which bring several improvements to the classical forward-backward methods. Our approach relies on the time discretization of a second-order differential system with two potentials, recently introduced by Attouch-Maingé-Redont [11].

The forward-backward algorithm has a long history going back to the projected gradient method. It has been first introduced by Lions and Mercier [33], and Passty [37] in order to find a zero of the sum of two maximal monotone operators $A + B$. The operator A is a general maximal monotone operator, while B is a maximal monotone operator which is cocoercive, and hence Lipschitz continuous (e.g. a Lipschitz-continuous gradient operator). It is a splitting method which successively computes a forward (explicit) step with respect to the smooth operator B and a backward (implicit) step with respect to the nonsmooth operator A. An important number of contributions, dealing with various topics, have been devoted to the development of this flexible method: monotone Lipschitz operators B that are not necessarily cocoercive (like linear skew-symmetric operators) with application to Lagrangian methods [25, 14, 44], acceleration of the method based on Nesterov's approach [16, 34, 35], approximate data and errors [8, 41], coupling with approximation methods [10]. For a recent account on these methods one can consult [8, 14, 16, 29, 41] and the bibliography therein.

H being a real Hilbert space, this paper is concerned with the minimization of a convex function $\Theta : H \to \mathbb{R} \cup \{+\infty\}$ that can be decomposed as $\Theta = \Phi + \Psi$, where $\Phi : H \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous convex function, and $\Psi : H \to \mathbb{R}$ is a differentiable function, not necessarily convex, whose gradient is Lipschitz continuous with constant L_Ψ.

With the support of the French ANR grant ANR-08-BLAN-0294-03. J. Peypouquet was partly supported by FONDECYT grant 11090023, Basal Project CMM Universidad de Chile, Núcleo Milenio Información y Coordinación en Redes ICM/FIC P10-024F, and Anillo Project ACT-1106: Analysis of Control Problems & Applications.

The classical forward-backward algorithm computes a sequence (u_k) starting from a given $u_0 \in H$ and recursively using the rule

$$u_{k+1} = (I + \lambda\partial\Phi)^{-1}\left(u_k - \lambda\nabla\Psi(u_k)\right), \quad \text{for } k \geq 1, \tag{1}$$

where $\lambda > 0$ is the step size. It is known that if $\mathcal{S} = \text{Argmin}(\Theta) \neq \emptyset$ and Ψ is convex, any sequence generated by algorithm (1) converges weakly to an element of $\mathcal{S}$, provided

$$0 < \lambda < \frac{2}{L_\Psi} \tag{2}$$

(see, for instance, [14, Theorem 25.8]). In some cases, in particular for large systems, the Lipschitz constant L_Ψ can be large, making it necessary to take a small λ. Or else, its exact evaluation may be computationally costly. All this can slow down the convergence of the algorithm. The inertial forward-backward algorithm introduced in this paper allows us to relax condition (2) in some respect.

First observe that the forward-backward algorithm (1) can be rewritten as

$$\frac{u_{k+1} - u_k}{\lambda} + \partial\Phi(u_{k+1}) + \nabla\Psi(u_k) \ni 0.$$

Therefore, it has a natural interpretation as a time discretization (implicit with respect to $\partial\Phi$ and explicit with respect to $\nabla\Psi$) of the first-order differential inclusion

$$\dot{u}(t) + \partial\Phi(u(t)) + \nabla\Psi(u(t)) \ni 0. \tag{3}$$

Recall also that, if $\mathcal{S} \neq \emptyset$, Theorem 4 in [20] asserts that each trajectory of (3) converges weakly to an element of $\mathcal{S}$.

As it occurs with (1) and (3), there is often a close relationship between discrete-time systems and their continuous-time counterparts, insofar as asymptotic behavior is concerned, see [38]. Keeping this in mind, our method is inspired by the second-order differential system with two potentials and Hessian driven damping introduced in [11]:

$$\ddot{u}(t) + \alpha\dot{u}(t) + \beta\nabla^2\Phi(u(t))\dot{u}(t) + \nabla\Phi(u(t)) + \nabla\Psi(u(t)) = 0. \tag{4}$$

From many viewpoints, this system has a rich interpretation in control theory (geometrical damping of nonlinear oscillators), mechanics (nonelastic shocks), PDE's (wave equation), see [4, 11]. Insofar as optimization is concerned, it is derived from the continuous Newton method (see [12, 5])

$$\nabla^2\Phi(u(t))\dot{u}(t) + \nabla\Phi(u(t)) = 0$$

and more precisely from its damped version

$$\left[\alpha I + \nabla^2\Phi(u(t))\right]\dot{u}(t) + \nabla\Phi(u(t)) = 0.$$

This system, which aims at minimizing function Φ with the variable metric given by $\alpha I + \nabla^2\Phi(u(t))$, may be generalized to minimize $\Phi + \Psi$ with the same metric:

$$\left[\alpha I + \nabla^2\Phi(u(t))\right]\dot{u}(t) + \nabla\Phi(u(t)) + \nabla\Psi(u(t)) = 0.$$

Now adding the second-order (inertial) term $\ddot{u}(t)$ gives impetus to the system and yields (4).

It turns out that (4) can be equivalently written as the first-order system:

$$\begin{cases} \dot{u}(t) + \beta\nabla\Phi(u(t)) + au(t) - by(t) = 0 \\ \dot{y}(t) + \beta\nabla\Psi(u(t)) - au(t) + by(t) = 0, \end{cases}$$

where a and b are real numbers such that $a + b = \alpha$ and $\beta b = 1$. It is important to observe that, since this alternative formulation does not involve the Hessian of Φ, it makes sense for nonsmooth convex functions (with $\nabla\Phi$ replaced by $\partial\Phi$). Note also that it makes sense for any choice of a, b and β; and so, up to re-parameterization and change of variables, one can take $\beta = 1$ for simplicity.

More precisely, start with $(u_0, y_0) \in H \times H$. At the k-th iteration, given (u_k, y_k) compute u_{k+1} and then y_{k+1} using the following rule:

$$\text{(IFB)} \qquad \begin{cases} 0 & \in & \dfrac{u_{k+1} - u_k}{\lambda} + \partial\Phi(u_{k+1}) + au_k - by_k \\ 0 & = & \dfrac{y_{k+1} - y_k}{\lambda} + \nabla\Psi(u_{k+1}) - au_{k+1} + by_{k+1}, \end{cases}$$

where a, b and λ are positive constants (IFB stands for Inertial Forward-Backward). As we shall see, this algorithm keeps the physical interpretation of (4), which will allow us to find energy-type Liapunov functions.

From a computational point-of-view, in comparison with the classical forward-backward algorithm (1), (IFB) only involves the addition of some terms whose computation is essentially costless. In return, they render the system sufficiently stable to ensure convergence under more general assumptions on the step size λ, and allow some lack of convexity for the regular function Ψ.

Optimization algorithms involving inertial features have been first introduced in [1, 2, 3, 6, 39] in connection with the second-order differential system

$$\ddot{u}(t) + \alpha\dot{u}(t) + \nabla\Phi(u(t)) = 0. \tag{5}$$

Because of their mechanical interpretation, they are called heavy ball with friction methods. When dealing with a nonsmooth potential Φ, the continuous dynamics (5) may involve elastic shocks, which raises severe difficulties concerning the existence and uniqueness of its solution trajectories [9]. By contrast, the introduction of the Hessian driven damping term $\nabla^2\Phi(u(t))\dot{u}(t)$ in (4) induces damped nonelastic shocks, which makes it a well posed equation. This basic idea has been first analyzed in [4] in the case of a single potential Φ. When dealing with two potentials Φ and Ψ, the same idea has been developed in [11] by introducing the Hessian driven damping term only for the nonsmooth potential Φ, giving rise to (4).

The paper is organized as follows. In section 1, we give the standing assumptions, and provide some insight regarding the stepsize selection. Section 2 is devoted to the asymptotic analysis of the algorithm and the proof of the main convergence result, namely Theorem 1. In section 3 we present conditions that guarantee the strong convergence of the algorithm and we describe the behavior of the auxiliary sequence (y_k). In section 4.1, we particularize our analysis to the case $\Phi = \delta_C$, the indicator function of the convex set C, and the corresponding gradient-projection method. In the last subsection, we give some numerical illustrations of our results to image recovery.

1. Standing Hypotheses

Recall that, given $\Phi : H \to \mathbb{R} \cup \{+\infty\}$ a proper lower semicontinuous convex function, for any $u \in H$, $\text{prox}_\Phi(u)$ is the unique solution of the strongly convex minimization problem

$$\min \left\{ \Phi(v) + \frac{1}{2}\|v - u\|^2 : \ v \in H \right\}.$$

The proximity operator $\text{prox}_\Phi : H \to H$ is a contraction (a firmly contractive mapping indeed; see [14]). Let us first reformulate the (IFB) algorithm with the help of the proximity mapping:

$$\begin{cases} u_{k+1} & = & \text{prox}_{\lambda\Phi}\left((1 - \lambda a)u_k + \lambda b y_k\right) \\ y_{k+1} & = & \dfrac{1}{1 + \lambda b} y_k + \dfrac{\lambda a}{1 + \lambda b} u_{k+1} - \dfrac{\lambda}{1 + \lambda b} \nabla\Psi(u_{k+1}). \end{cases} \tag{6}$$

It is a first-order discrete dynamics with respect to the couple (u_k, y_k): at the k-th iteration, given (u_k, y_k), first compute u_{k+1} with the help of the proximity mapping of Φ (implicit, backward step), and then y_{k+1}, by using the gradient of Ψ (explicit, forward step).

Most of the asymptotic analysis of the algorithm given by (IFB) relies on the following set of hypotheses:

Hypothesis H:

- $\mathbf{H}_\Phi$: The function $\Phi : H \to \mathbf{R} \cup \{+\infty\}$ is proper, lower-semicontinuous and convex.
- $\mathbf{H}_\Psi$: The function $\Psi : H \to \mathbf{R}$ is differentiable and its gradient $\nabla\Psi$ is Lipschitz continuous with constant L_Ψ.
- $\mathbf{H}_\lambda$: The step size λ satisfies $0 < \lambda < \Lambda = \min\left\{\frac{1}{a}, \frac{2(a+b)}{bL_\Psi}\right\}$, with $a, b > 0$.
- $\mathbf{H}_\Theta$: The function $\Theta = \Phi + \Psi$ is convex and bounded from below.

A few comments regarding the selection of the step size λ are in order:

First, it is not particularly interesting to take $a \geq \frac{L_\Psi}{2}$ because in that case $\Lambda \leq \frac{2}{L_\Psi}$ and there is no improvement with respect to the classical bound for the step size. Also, a large value of a would probably introduce unwanted numerical instabilities.

Next, whenever $0 < a < \frac{L_\Psi}{2}$, by choosing $b \leq \frac{2a^2}{L_\Psi - 2a}$, we obtain $\Lambda = \frac{1}{a}$. Therefore, Λ can be made arbitrarily large by choosing a reasonably small. In particular, the naive choice $a = b = \frac{L_\Psi}{4}$ gives $\Lambda = \frac{4}{L_\Psi}$, which is already twice as large as the classical $\frac{2}{L_\Psi}$.

Finally, when L_Ψ is large, it would be best to be able to tune all the parameters a, b and λ in such a way that none of them is either too large or too small. One way to do this is to take $a = \frac{1}{2}L_\Psi^{\frac{1}{N}}$ and $b = \frac{1}{2}L_\Psi^{\frac{2}{N}-1}$ with $N \geq 2$ (but not too large), so that $\Lambda = 2L_\Psi^{-\frac{1}{N}}$. A logarithmic rule may also do the trick.

2. Asymptotic analysis of the algorithm

The main purpose of this section is to prove the following:

Theorem 1. *Let Hypothesis* **H** *hold. Then* $\lim_{k\to+\infty} \Theta(u_k) = \inf \Theta$.
Assume moreover that $\mathcal{S} = \text{Argmin}\, \Theta \neq \emptyset$ *and any of the following conditions holds:*

i) *The set* $\mathcal{S}$ *is a singleton;*
ii) *The function* Φ *is differentiable and its gradient is* weak-to-weak *sequentially continuous;*
iii) *The gradient of* Ψ *is* weak-to-weak *sequentially continuous;*
iv) *The function* Ψ *is convex.*

Then any sequence (u_k) *generated by algorithm* (IFB) *converges weakly as* $k \to +\infty$ *to an element of* $\mathcal{S}$. *On the other hand, if* $\mathcal{S} = \emptyset$, *then* $\lim_{k\to+\infty} \|u_k\| = +\infty$.

Since the argument is rather technical, in order to simplify the presentation of the proof, we begin by describing the strategy: First, we establish that an energy-type sequence, denoted below by (E_k), is decreasing, thus convergent. Next, for each $q \in H$, we define a sequence, denoted by $(F_k(q))$, that exhibits a Liapunov-type property whenever $q \in \mathcal{S}$. As we shall see, these dissipation properties will allow us to deduce that (u_k) is a minimizing sequence for Θ, and thus every weak cluster point of (u_k) belongs to $\mathcal{S}$. Finally, we conclude by showing that if $\mathcal{S} \neq \emptyset$, then (u_k) is bounded and has exactly one weak cluster point.

2.1. Energy decrease and other dissipative properties.

In order to simplify the notation define

$$\gamma = \frac{1 - a\lambda}{2b\lambda^2} > 0, \tag{7}$$

where the inequality follows from $\mathbf{H}_\lambda$. The velocity of the system is represented by the sequence (ξ_k) defined by

$$\xi_k = u_k - u_{k-1}.$$

In this section we prove that the mechanical energy (E_k) of the system, defined by

$$E_k = \Theta(u_k) + \gamma\|\xi_k\|^2, \tag{8}$$

is decreasing; that the sequence of velocities belongs to ℓ^2; and that the sequence of values $(\Theta(u_k))$ converges, even if it is not necessarily decreasing. In order to simplify the notation write

$$\zeta_k = y_k - y_{k-1} \tag{9}$$

and

$$z_k = \frac{\xi_k}{\lambda} + au_{k-1} - by_{k-1} \in -\partial\Phi(u_k). \tag{10}$$

The main result of this subsection is:

Proposition 2. *Let Hypothesis* **H** *hold* [1]. *Then the sequence* (E_k) *is nonincreasing and* $\lim_{k\to+\infty} E_k$ *exists. Moreover,* $(\|\xi_k\|) \in \ell^2$ *and* $\lim_{k\to+\infty} \Theta(u_k) = \lim_{k\to+\infty} E_k$.

Proof. Since Θ is bounded from below, so is (E_k). Therefore, it suffices to prove that there exists $\beta > 0$ such that $E_{k+1} - E_k + \beta\|\xi_{k+1}\|^2 \leq 0$ for every $k \geq 1$.

First, let us write the subdifferential inequality for Φ at point u_{k+1}; for all $v \in H$, we have

$$\Phi(v) \geq \Phi(u_{k+1}) + \langle \frac{u_{k+1} - u_k}{\lambda} + au_k - by_k, u_{k+1} - v\rangle.$$

Setting $v = u_k$, we deduce that

$$\Phi(u_k) \geq \Phi(u_{k+1}) + \langle \frac{u_{k+1} - u_k}{\lambda} + au_k - by_k, u_{k+1} - u_k\rangle \ ,$$

and so

$$\Phi(u_{k+1}) - \Phi(u_k) + \frac{1}{\lambda}\|\xi_{k+1}\|^2 + \langle au_k - by_k, \xi_{k+1}\rangle \leq 0. \tag{11}$$

On the other hand, the Descent Lemma for Ψ ([17, Prop. A.24], [36]) gives

$$\Psi(v) \leq \Psi(u_k) + \langle \nabla\Psi(u_k), v - u_k\rangle + \frac{L_\Psi}{2}\|v - u_k\|^2,$$

which, in view of the second (IFB) equation, yields

$$\Psi(v) \leq \Psi(u_k) + \langle au_k - by_k - \frac{y_k - y_{k-1}}{\lambda}, v - u_k\rangle + \frac{L_\Psi}{2}\|v - u_k\|^2.$$

Setting $v = u_{k+1}$, we obtain

$$\Psi(u_{k+1}) - \Psi(u_k) - \frac{L_\Psi}{2}\|\xi_{k+1}\|^2 - \langle au_k - by_k, \xi_{k+1}\rangle \leq -\frac{1}{\lambda}\langle \zeta_k, \xi_{k+1}\rangle, \tag{12}$$

where ζ_k is defined in (9). Adding inequalities (11) and (12) we obtain

$$\Theta(u_{k+1}) - \Theta(u_k) + \left[\frac{1}{\lambda} - \frac{L_\Psi}{2}\right]\|\xi_{k+1}\|^2 \leq -\frac{1}{\lambda}\langle \zeta_k, \xi_{k+1}\rangle. \tag{13}$$

Recall from (10) that $z_{k+1} = \frac{\xi_{k+1}}{\lambda} + au_k - by_k \in -\partial\Phi(u_{k+1})$. We use the monotonicity of $\partial\Phi$ at points u_k and u_{k+1} to deduce that

$$\langle \frac{\xi_{k+1}}{\lambda} + au_k - by_k - \frac{\xi_k}{\lambda} - au_{k-1} + by_{k-1}, u_{k+1} - u_k\rangle \leq 0.$$

We can rewrite this as

$$\frac{1}{\lambda}\langle \xi_{k+1} - \xi_k, \xi_{k+1}\rangle + a\langle \xi_k, \xi_{k+1}\rangle - b\langle \zeta_k, \xi_{k+1}\rangle \leq 0.$$

Dividing by $b\lambda$ and rearranging the terms we obtain, with γ defined by (7)

$$-\frac{1}{\lambda}\langle \zeta_k, \xi_{k+1}\rangle + \frac{1}{b\lambda^2}\|\xi_{k+1}\|^2 - 2\gamma\langle \xi_k, \xi_{k+1}\rangle \leq 0. \tag{14}$$

[1] actually, the convexity of Θ is not used here.

Since

$$2\langle \xi_k, \xi_{k+1}\rangle \leq \|\xi_{k+1}\|^2 + \|\xi_k\|^2,$$

inequality (14) gives

$$-\frac{1}{\lambda}\langle \zeta_k, \xi_{k+1}\rangle + \left[\frac{1}{b\lambda^2} - \gamma\right] \|\xi_{k+1}\|^2 - \gamma\|\xi_k\|^2 \leq 0. \tag{15}$$

Adding inequalities (13) and (15) we see that

$$\Theta(u_{k+1}) - \Theta(u_k) + \left[\frac{1}{\lambda} - \frac{L_\Psi}{2} + \frac{1}{b\lambda^2} - \gamma\right] \|\xi_{k+1}\|^2 - \gamma\|\xi_k\|^2 \leq 0$$

and recalling the definition of E_k we deduce that

$$E_{k+1} - E_k + \left[\frac{1}{\lambda} - \frac{L_\Psi}{2} + \frac{1}{b\lambda^2} - 2\gamma\right] \|\xi_{k+1}\|^2 \leq 0.$$

Having $\mathbf{H}_\lambda$ in mind, it suffices to set

$$\beta = \frac{1}{\lambda} - \frac{L_\Psi}{2} + \frac{1}{b\lambda^2} - 2\gamma = \frac{2(a+b) - b\lambda L_\Psi}{2b\lambda} > 0 \tag{16}$$

to complete the proof. ■

2.2. **A Liapunov-type sequence.** The technical issues are gathered in this subsection. Given $q \in H$ and $k \geq 2$, we define the auxiliary sequence $(G_k(q))$ by

$$G_k(q) = \langle z_k, u_k - q\rangle - \sum_{i=2}^{k} \langle z_i, \xi_i\rangle, \tag{17}$$

where z_k is given by (10). Next, we define the sequence $(F_k(q))$, by

$$F_k(q) = \frac{a+b}{2\lambda}\|u_k - q\|^2 - \left(\frac{1}{\lambda} + b\right) G_{k+1}(q) - \delta \sum_{i=2}^{k} \|\xi_i\|^2 + 2\alpha\langle \xi_{k+1}, u_k - q\rangle, \tag{18}$$

with

$$\alpha = \frac{1 + b\lambda}{2\lambda^2} \quad \text{and} \quad \delta = \frac{2 + (b-a)\lambda}{2\lambda^2}.$$

Sequence $(F_k(q))$ will play a central role in the convergence analysis because of its close relationship with the function Θ, given in the following:

Proposition 3. *Assume Θ is convex. For each $q \in H$ and $k \geq 1$ we have*

$$b(\Theta(q) - \Theta(u_{k+1})) \geq b\left\langle \left(\frac{1}{\lambda} - a\right) \xi_{k+1} + \left(\frac{1}{\lambda} + b\right) \zeta_{k+1}, u_{k+1} - q \right\rangle = F_{k+1}(q) - F_k(q). \tag{19}$$

In particular, the sequence $(F_k(q))$ is nonincreasing whenever $q \in \mathcal{S}$.

Proof. First observe that, since Θ is convex and

$$0 \in \left(\frac{1}{\lambda} - a\right) \xi_k + \left(\frac{1}{\lambda} + b\right) \zeta_k + \partial\Theta(u_k), \tag{20}$$

we must have

$$\left\langle \left(\frac{1}{\lambda} - a\right) \xi_k + \left(\frac{1}{\lambda} + b\right) \zeta_k, u_k - q \right\rangle \leq \Theta(q) - \Theta(u_k),$$

which proves the inequality. Let us now prove that the product above can be written as the difference of consecutive terms of the sequence $(F_k(q))$. From the definition of $G_k(q)$ given in (17) we have

$$G_{k+2}(q) - G_{k+1}(q) = \langle z_{k+2}, u_{k+2} - q\rangle - \langle z_{k+1}, u_{k+1} - q\rangle - \langle z_{k+2}, \xi_{k+2}\rangle.$$

We deduce that

$$\begin{aligned} F_{k+1}(q) - F_k(q) &= \frac{a+b}{2\lambda}\left[\|u_{k+1} - q\|^2 - \|u_k - q\|^2\right] + \left(\frac{1}{\lambda} + b\right)\langle z_{k+2}, \xi_{k+2}\rangle - \delta\|\xi_{k+1}\|^2 \\ &\quad - \left(\frac{1}{\lambda} + b\right)\left[\langle z_{k+2}, u_{k+2} - q\rangle - \langle z_{k+1}, u_{k+1} - q\rangle\right] \\ &\quad + 2\alpha\left[\langle \xi_{k+2}, u_{k+1} - q\rangle - \langle \xi_{k+1}, u_k - q\rangle\right] \\ &= \frac{a+b}{2\lambda}\left[\|u_{k+1} - q\|^2 - \|u_k - q\|^2\right] + \left(\frac{1}{\lambda} + b\right)\langle z_{k+2}, \xi_{k+2}\rangle - \delta\|\xi_{k+1}\|^2 \\ &\quad - \left(\frac{1}{\lambda} + b\right)\left[\langle z_{k+2}, u_{k+1} - q\rangle - \langle z_{k+1}, u_k - q\rangle\right] \\ &\quad + 2\alpha\left[\langle \xi_{k+2}, u_{k+1} - q\rangle - \langle \xi_{k+1}, u_k - q\rangle\right] \\ &\quad - \left(\frac{1}{\lambda} + b\right)\left[\langle z_{k+2}, \xi_{k+2}\rangle - \langle z_{k+1}, \xi_{k+1}\rangle\right] \\ &= \frac{a+b}{2\lambda}\left[\|u_{k+1} - q\|^2 - \|u_k - q\|^2\right] + \left(\frac{1}{\lambda} + b\right)\langle z_{k+1}, \xi_{k+1}\rangle - \delta\|\xi_{k+1}\|^2 \\ &\quad - \left(\frac{1}{\lambda} + b\right)\left[\langle z_{k+2} - \frac{\xi_{k+2}}{\lambda}, u_{k+1} - q\rangle - \langle z_{k+1} - \frac{\xi_{k+1}}{\lambda}, u_k - q\rangle\right] \\ &= \frac{a+b}{2\lambda}\left[\|u_{k+1} - q\|^2 - \|u_k - q\|^2\right] + \left(\frac{1}{\lambda} + b\right)\langle z_{k+1}, \xi_{k+1}\rangle - \delta\|\xi_{k+1}\|^2 \\ &\quad - \left(\frac{1}{\lambda} + b\right)\left[\langle a u_{k+1} - b y_{k+1}, u_{k+1} - q\rangle - \langle a u_k - b y_k, u_k - q\rangle\right]. \end{aligned} \tag{21}$$

Keeping this in mind, let us rewrite the quantity

$$B_{k+1}(q) = b\left\langle \left(\frac{1}{\lambda} - a\right)\xi_{k+1} + \left(\frac{1}{\lambda} + b\right)\zeta_{k+1}, u_{k+1} - q\right\rangle$$

on the right-hand side of the equality in (19). First observe that

$$\begin{aligned} B_{k+1}(q) &= \frac{b}{\lambda}\langle \xi_{k+1}, u_{k+1} - q\rangle - b\langle a\xi_{k+1}, u_{k+1} - q\rangle + \left(\frac{1}{\lambda} + b\right)\langle b\zeta_{k+1}, u_{k+1} - q\rangle \\ &= \frac{a+b}{\lambda}\langle \xi_{k+1}, u_{k+1} - q\rangle - \left(\frac{1}{\lambda} + b\right)\langle a\xi_{k+1}, u_{k+1} - q\rangle + \left(\frac{1}{\lambda} + b\right)\langle b\zeta_{k+1}, u_{k+1} - q\rangle \\ &= \frac{a+b}{\lambda}\langle \xi_{k+1}, u_{k+1} - q\rangle - \left(\frac{1}{\lambda} + b\right)\langle a\xi_{k+1} - b\zeta_{k+1}, u_{k+1} - q\rangle \\ &= \frac{a+b}{2\lambda}\left[\|u_{k+1} - q\|^2 - \|u_k - q\|^2 + \|\xi_{k+1}\|^2\right] - \left(\frac{1}{\lambda} + b\right)\langle a\xi_{k+1} - b\zeta_{k+1}, u_{k+1} - q\rangle. \end{aligned} \tag{22}$$

Next, we have

$$\begin{aligned}\langle a\xi_{k+1}-b\zeta_{k+1},u_{k+1}-q\rangle &= \langle au_{k+1}-by_{k+1},u_{k+1}-q\rangle-\langle au_k-by_k,u_k-q\rangle-\langle au_k-by_k,\xi_{k+1}\rangle\\ &= \langle au_{k+1}-by_{k+1},u_{k+1}-q\rangle-\langle au_k-by_k,u_k-q\rangle+\frac{1}{\lambda}\|\xi_{k+1}\|^2\\ &\quad -\langle\frac{\xi_{k+1}}{\lambda}+au_k-by_k,\xi_{k+1}\rangle\\ &= \langle au_{k+1}-by_{k+1},u_{k+1}-q\rangle-\langle au_k-by_k,u_k-q\rangle\\ &\quad +\frac{1}{\lambda}\|\xi_{k+1}\|^2-\langle z_{k+1},\xi_{k+1}\rangle.\end{aligned}$$

Using this in (22) we deduce that

$$\begin{aligned}B_{k+1}(q) &= \frac{a+b}{2\lambda}\left[\|u_{k+1}-q\|^2-\|u_k-q\|^2\right]+\left[\frac{a+b}{2\lambda}-\frac{1}{\lambda}\left(\frac{1}{\lambda}+b\right)\right]\|\xi_{k+1}\|^2\\ &\quad -\left(\frac{1}{\lambda}+b\right)\left[\langle au_{k+1}-by_{k+1},u_{k+1}-q\rangle-\langle au_k-by_k,u_k-q\rangle\right]+\left(\frac{1}{\lambda}+b\right)\langle z_{k+1},\xi_{k+1}\rangle\\ &= \frac{a+b}{2\lambda}\left[\|u_{k+1}-q\|^2-\|u_k-q\|^2\right]+\left(\frac{1}{\lambda}+b\right)\langle z_{k+1},\xi_{k+1}\rangle-\delta\|\xi_{k+1}\|^2\\ &\quad -\left(\frac{1}{\lambda}+b\right)\left[\langle au_{k+1}-by_{k+1},u_{k+1}-q\rangle-\langle au_k-by_k,u_k-q\rangle\right].\end{aligned} \tag{23}$$

The right-hand sides of (21) and (23) coincide, and the equality in (19) is verified. ■

An important property of the sequence $(G_k(q))$ is the following:

Proposition 4. *Let Hypotheses* $\mathbf{H}_\Phi$ *and* $\mathbf{H}_\Psi$ *hold. Then, for each* $q\in dom(\Phi)$*, the sequence* $(G_k(q))$ *is bounded from above.*

Proof. First write

$$G_k(q)=\langle z_k,u_k-q\rangle-\sum_{i=2}^{k}\langle z_{i-1},u_i-u_{i-1}\rangle-\sum_{i=2}^{k}\langle z_i-z_{i-1},\xi_i\rangle. \tag{24}$$

The subdifferential inequality for Φ gives

$$\langle z_k,u_k-q\rangle\le\Phi(q)-\Phi(u_k) \tag{25}$$

and

$$-\langle z_{i-1},u_i-u_{i-1}\rangle\le\Phi(u_i)-\Phi(u_{i-1}) \tag{26}$$

because $-z_j\in\partial\Phi(u_j)$ for each j. On the other hand,

$$\begin{aligned}-\langle z_i-z_{i-1},\xi_i\rangle &= -\langle\frac{\xi_i-\xi_{i-1}}{\lambda}+a\xi_{i-1}-b\zeta_{i-1},\xi_i\rangle\\ &= -\frac{1}{\lambda}\|\xi_i\|^2+\frac{1-a\lambda}{\lambda}\langle\xi_{i-1},\xi_i\rangle+b\langle\zeta_{i-1},\xi_i\rangle\\ &\le -\frac{1}{\lambda}\|\xi_i\|^2+\frac{1-a\lambda}{2\lambda}\|\xi_i\|^2+\frac{1-a\lambda}{2\lambda}\|\xi_{i-1}\|^2\\ &\quad +b\lambda(\Theta(u_{i-1})-\Theta(u_i))-b\lambda\left[\frac{1}{\lambda}-\frac{L_\Psi}{2}\right]\|\xi_i\|^2\\ &\le \left[\frac{b\lambda^2L_\Psi-(a+2b)\lambda-1}{2\lambda}\right]\|\xi_i\|^2+\frac{1-a\lambda}{2\lambda}\|\xi_{i-1}\|^2+b\lambda(\Theta(u_{i-1})-\Theta(u_i))\end{aligned} \tag{27}$$

by inequality (13). Using inequalities (25), (26) and (27) in equality (24) we deduce that

$$G_k(q) \le \Phi(q) - \Phi(u_1) + b\lambda(\Theta(u_1) - \Theta(u_k)) + K \sum_{i=1}^{k} \|\xi_i\|^2$$

for some positive constant K. The right-hand side is finite by virtue of Proposition 2. ■

2.3. **Minimization, boundedness and stability.** We begin by observing that, by replacing

$$2\langle \xi_{k+1}, u_k - q\rangle = \|u_{k+1} - q\|^2 - \|u_k - q\|^2 - \|\xi_{k+1}\|^2$$

in the definition of $F_k(q)$ given in (18), we easily deduce from Proposition 2 and Proposition 4 that

$$F_k(q) \ge C + \alpha\|u_{k+1} - q\|^2 - \left[\alpha - \frac{a+b}{2\lambda}\right] \|u_k - q\|^2 \ge C + \alpha\left[\|u_{k+1} - q\|^2 - \|u_k - q\|^2\right] \tag{28}$$

for some constant C. This double inequality will be useful to prove Propositions 6 and 8.

In the first place, we show that the sequence (u_k) minimizes Θ, even if $\mathcal{S} = \emptyset$. We shall use the following result concerning real sequences:

Lemma 5. *Let (a_k), (b_k) and (ε_k) be real sequences with $(\varepsilon_k) \in \ell^1$. Let $\theta \in \mathbf{R}$ and assume*

$$a_k - a_{k-1} \le \theta + \varepsilon_k, \qquad b_{k+1} - b_k \le a_k \quad \text{and} \quad b_k \ge 0$$

for all $k \ge k_0$. Then $\theta \ge 0$.

Proof. Summing up the first inequality for $k = k_0, \ldots, m$ we deduce that $a_m \le m\theta + A$ for some constant A. Using the second inequality and summing again for $m = k_0, \ldots, M$ we obtain

$$b_{M+1} - b_{k_0} \le \sum_{k=k_0}^{M} a_k \le \sum_{k=1}^{M} (k\theta + A) = \frac{M(M+1)}{2}\theta + MA.$$

Thus $0 \le 2b_{M+1} \le M^2\theta + (2A + \theta)M + 2b_{k_0}$. Divide by M^2 and let $M \to +\infty$ to see that $\theta \ge 0$. ■

Now we are in a position to prove:

Proposition 6. *Assume hypothesis* **H** *holds. Then* $\lim_{k\to+\infty} \Theta(u_k) = \inf \Theta$. *In particular, every weak cluster point of the sequence (u_k) lies in $\mathcal{S}$. As a consequence, if $\mathcal{S} = \emptyset$ then* $\lim_{k\to+\infty} \|u_k\| = +\infty$.

Proof. From Proposition 3 and the definition of E_k we deduce that, for each $q \in H$,

$$F_k(q) - F_{k-1}(q) \le b(\Theta(q) - \Theta(u_k)) = b(\Theta(q) - E_k) + b\gamma\|\xi_k\|^2.$$

Let $E_\infty = \lim_{k\to+\infty} E_k = \lim_{k\to+\infty} \Theta(u_k)$. Since the sequence (E_k) is nonincreasing we have

$$F_k(q) - F_{k-1}(q) \le b(\Theta(q) - E_\infty) + b\gamma\|\xi_k\|^2.$$

On the other hand, inequality (28) shows that

$$\alpha\|u_{k+1} - q\|^2 - \alpha\|u_k - q\|^2 \le F_k(q) - C$$

for some constant C. Taking $a_k = F_k(q) - C$, $b_k = \alpha\|u_k - q\|^2$, $\varepsilon_k = b\gamma\|\xi_k\|^2$ (which is in ℓ^1 by Proposition 2) and $\theta = b(\Theta(q) - E_\infty)$ in Lemma 5, we deduce that $E_\infty \le \Theta(q)$. Since this holds for any $q \in H$ we have $E_\infty = \inf \Theta$. ■

Observe that the sequence (u_k) is minimizing even if $\mathcal{S} = \emptyset$. Let us now study the case $\mathcal{S} \ne \emptyset$. We require a simple auxiliary result:

Lemma 7. *Let (p_k) be a nonnegative sequence such that*

$$p_k - \omega p_{k-1} \le K$$

for some $\omega < 1$, some $K \in \mathbf{R}$ and all sufficiently large k. Then (p_k) is bounded.

Proof. If $\omega \le 0$ the result is immediate. If $\omega \in (0,1)$, a simple induction argument shows that $p_k \le \omega^k p_0 + K \sum_{j=0}^{k-1} \omega^j \le p_0 + K(1-\omega)^{-1}$ and so (p_k) is bounded. ■

Proposition 8. *Assume Hypothesis **H** holds and let $\mathcal{S} \neq \emptyset$. Then the sequence (u_k) is bounded. Moreover, $\lim_{k\to+\infty} F_k(q)$ exists for each $q \in \mathcal{S}$.*

Proof. Take $q \in \mathcal{S}$. By Proposition 3, the sequence $(F_k(q))$ is nonincreasing. Inequality (28) then gives

$$\|u_k - q\|^2 - \left[\frac{2\alpha\lambda - a - b}{2\alpha\lambda}\right] \|u_{k-1} - q\|^2 \le \left[\frac{F_1(q) - C}{\alpha}\right].$$

Since $\frac{2\alpha\lambda - a - b}{2\alpha\lambda} \in (0,1)$, Lemma 7 shows that the sequence $(\|u_k - q\|)$ is bounded. Using inequality (28) again we deduce that the sequence $(F_k(q))$ is bounded from below. Since it is nonincreasing, it must converge. ■

2.4. **Proof of Theorem 1.** The final statement follows directly from Proposition 6. Let us assume that $\mathcal{S} \neq \emptyset$, and prove the weak convergence of (u_k) to an element of $\mathcal{S}$. We shall analyze each case separately:

Case i**):** It is immediate from Propositions 6 and 8.

For cases ii) and iii) we shall prove that the sequence (u_k) can have at most one weak cluster point. But before doing so, let us turn our attention to the first two terms in the definition of $F_k(q)$. For $q \in H$ define

$$A_k(q) = \left(\frac{a+b}{2\lambda}\right) \|u_k - q\|^2 - \left(\frac{1}{\lambda} + b\right) G_{k+1}(q) \tag{29}$$

and observe that $\lim_{k\to+\infty} A_k(q)$ exists whenever $q \in \mathcal{S}$. Indeed, recall from (18) that

$$F_k(q) = A_k(q) - \delta \sum_{i=2}^{k} \|\xi_i\|^2 + 2\alpha\langle \xi_{k+1}, u_k - q\rangle.$$

We know that $(\|\xi_i\|)$ belongs to ℓ^2 by Proposition 2 and that (u_k) is bounded by Proposition 8; hence $\lim_{k\to+\infty} F_k(q) - A_k(q)$ exists. But $\lim_{k\to+\infty} F_k(q)$ also exists since $q \in \mathcal{S}$ (also by Proposition 8).

Now take $p, q \in H$ and $m, n \in \mathbf{N}$. We have

$$A_m(p) - A_n(p) - A_m(q) + A_n(q) = \left(\frac{a+b}{\lambda}\right) \langle u_m - u_n, q - p\rangle - \left(\frac{1}{\lambda} + b\right) \langle z_{m+1} - z_{n+1}, q - p\rangle. \tag{30}$$

Assume $u_{t_k} \rightharpoonup \bar{p}$ and $u_{s_k} \rightharpoonup \bar{q}$ as $k \to +\infty$. By Proposition 6, both $\bar{p}$ and $\bar{q}$ must be in $\mathcal{S}$. Observe also that, since $\lim_{k\to+\infty} \xi_k = 0$ we have $u_{t_k - 1} \rightharpoonup \bar{p}$ and $u_{s_k - 1} \rightharpoonup \bar{q}$ as $k \to +\infty$. Use $p = \bar{p}$, $q = \bar{q}$, $m = t_k - 1$ and $n = s_k - 1$ in (30). Then let $k \to +\infty$ to obtain

$$\lim_{k\to+\infty} \langle z_{t_k} - z_{s_k}, \bar{p} - \bar{q}\rangle = \left(\frac{a+b}{1+b\lambda}\right) \|\bar{p} - \bar{q}\|^2. \tag{31}$$

This fact will be used in the sequel.

Case *ii*)**:** Since $\nabla\Phi$ is weak-to-weak sequentially continuous, if $u_{t_k} \rightharpoonup \bar{p}$ and $u_{s_k} \rightharpoonup \bar{q}$ as $k \to +\infty$, we deduce that

$$\lim_{k\to+\infty} \langle z_{t_k} - z_{s_k}, \bar{p} - \bar{q}\rangle = -\langle \nabla\Phi(\bar{p}) - \nabla\Phi(\bar{q}), \bar{p} - \bar{q}\rangle \le 0$$

by the monotonicity of $\nabla\Phi$. Then (31) implies that $\bar{p} = \bar{q}$.

Case *iii*)**:** Observe that $\left(\frac{1}{\lambda} - a\right)\xi_k + \left(\frac{1}{\lambda} + b\right)\zeta_k = z_k - \nabla\Psi(u_k)$. Proposition 3 implies

$$F_k(\bar{q}) - F_{k-1}(\bar{q}) = b\langle z_k - \nabla\Psi(u_k), u_k - \bar{q}\rangle \tag{32}$$

and this quantity vanishes as $k \to +\infty$ whenever $\bar{q} \in \mathcal{S}$. For $q \in H$ set $\mathcal{A}_k(q) = \langle z_k - \nabla\Psi(u_k), u_k - q\rangle$. A simple computation shows that for $p, q \in H$ and for $m, n \in \mathbf{N}$ we have

$$\langle \nabla\Psi(u_m) - \nabla\Psi(u_n), p - q\rangle = \mathcal{A}_m(p) - \mathcal{A}_n(p) - \mathcal{A}_m(q) + \mathcal{A}_n(q) + \langle z_m - z_n, p - q\rangle.$$

If $u_{t_k} \rightharpoonup \bar{p}$ and $u_{s_k} \rightharpoonup \bar{q}$ as $k \to +\infty$, then $\bar{p}, \bar{q} \in \mathcal{S}$ and the previous equation gives

$$\lim_{k\to\infty} \langle \nabla\Psi(u_{t_k}) - \nabla\Psi(u_{s_k}), \bar{p} - \bar{q}\rangle = \lim_{k\to\infty} \langle z_{t_k} - z_{s_k}, \bar{p} - \bar{q}\rangle$$

Whence, in view of (31) and the weak-to-weak sequential continuity of $\nabla\Psi$, we conclude that

$$\left(\frac{a+b}{1+b\lambda}\right)\|\bar{p} - \bar{q}\|^2 = \langle \nabla\Psi(\bar{p}) - \nabla\Psi(\bar{q}), \bar{p} - \bar{q}\rangle \le 0$$

because $\nabla\Psi(q) \in -\partial\Phi(q)$ whenever $q \in \mathcal{S}$. It ensues, as before, that $\bar{p} = \bar{q}$.

Case *iv*)**:** Let $q \in \mathcal{S}$ and $m > n$. We have

$$G_{n+1}(q) - G_{m+1}(q) = \left[\langle z_{n+1}, u_{n+1} - q\rangle + \sum_{i=n+2}^{m+1} \langle z_i, \xi_i\rangle\right] + \langle -z_{m+1}, u_{m+1} - q\rangle.$$

On the one hand,

$$\begin{aligned}
\langle z_{n+1}, u_{n+1} - q\rangle + \sum_{i=n+2}^{m+1} \langle z_i, \xi_i\rangle &\le \Phi(q) - \Phi(u_{n+1}) + \sum_{i=n+2}^{m+1} [\Phi(u_{i-1}) - \Phi(u_i)] \\
&= \Phi(q) - \Phi(u_{m+1})
\end{aligned}$$

by the convexity of Φ. On the other hand, equality (32) and the convexity of Ψ imply

$$\begin{aligned}
\langle -z_{m+1}, u_{m+1} - q\rangle &= \langle -z_{m+1} + \nabla\Psi(u_{m+1}), u_{m+1} - q\rangle - \langle \nabla\Psi(u_{m+1}), u_{m+1} - q\rangle \\
&\le \frac{1}{b}[F_m(q) - F_{m+1}(q)] + \Psi(q) - \Psi(u_{m+1}).
\end{aligned}$$

From the above relations and the definition of G, it follows that

$$G_{n+1}(q) - G_{m+1}(q) \le \Theta(q) - \Theta(u_{m+1}) + \frac{1}{b}[F_m(q) - F_{m+1}(q)].$$

Whence, in view of (29),

$$\begin{aligned}
\left(\frac{a+b}{2\lambda}\right)\|u_n - q\|^2 &= \left(\frac{a+b}{2\lambda}\right)\|u_m - q\|^2 + [A_n(q) - A_m(q)] + \left(\frac{1}{\lambda} + b\right)[G_{n+1}(q) - G_{m+1}(q)] \\
&\le \left(\frac{a+b}{2\lambda}\right)\|u_m - q\|^2 + [A_n(q) - A_m(q)] \\
&\quad + \left(\frac{1}{\lambda} + b\right)\left[\frac{1}{b}[F_m(q) - F_{m+1}(q)] + \Theta(q) - \Theta(u_{m+1})\right].
\end{aligned}$$

Take the lower limit as $m \to +\infty$ and recall that $\lim_{m\to+\infty} \Theta(u_m) = \Theta(q)$. Then take the upper limit as $n \to +\infty$ to conclude that

$$\limsup_{n\to+\infty} \|u_n - q\| \le \liminf_{m\to+\infty} \|u_m - q\|.$$

Therefore $\lim_{k\to+\infty}\|u_k - q\|$ exists and the result follows from Proposition 6 and Opial's Lemma. ■

3. Further convergence properties

3.1. **Strong convergence.** Let us present some conditions that guarantee the strong convergence of the sequence (u_k).

3.1.1. *Finite dimension.* If H is finite-dimensional, condition iii) in Theorem 1 automatically holds under Hypothesis $\mathbf{H}_\Psi$. Therefore, no supplementary conditions are needed and we have the following:

Corollary 9. *Let H be finite-dimensional. Assume Hypothesis* $\mathbf{H}$ *holds and $\mathcal{S}\neq\emptyset$. Then (u_k) converges to an element of $\mathcal{S}$.*

3.1.2. *Strong convexity.* Assume Θ is strongly convex with parameter ρ. In view of (20), for each $q\in H$ and $k\geq 1$ we have

$$\left\langle \left(\frac{1}{\lambda}-a\right)\xi_k + \left(\frac{1}{\lambda}+b\right)\zeta_k, u_k - q\right\rangle + \frac{\rho}{2}\|u_k-q\|^2 \leq \Theta(q)-\Theta(u_k),$$

and so

$$F_{k+1}(q)-F_k(q)+\frac{\rho\, b}{2}\|u_{k+1}-q\|^2 \leq b(\Theta(q)-\Theta(u_{k+1})), \tag{33}$$

by the equality in (19). Moreover, since Θ is strongly convex, we have $\mathcal{S}=\{\bar q\}$. If Hypothesis $\mathbf{H}$ holds, we can write $q=\bar q$ in inequality (33), let $k\to+\infty$ and apply Propositions 6 and 8 to deduce that $\lim_{k\to+\infty}\|u_k-\bar q\|=0$. We have proved the following:

Corollary 10. *Assume Hypothesis* $\mathbf{H}$ *holds and let Θ be strongly convex. Then (u_k) converges strongly to the unique $\bar q\in\mathcal{S}$.*

3.1.3. *Bounded inf-compactness.* Here we present another condition that ensures that the convergence is actually strong. A function $f : H\to\mathbf{R}\cup\{+\infty\}$ is *boundedly inf-compact* if for each $r\in\mathbf{R}$ and $R>0$ the set

$$\Gamma_r^R = \{\, v\in H \;:\; f(v)\leq r \text{ and } \|v\|\leq R\,\}$$

is relatively compact for the strong topology. We have the following:

Corollary 11. *Assume Hypothesis* $\mathbf{H}$ *holds and let Θ be boundedly inf-compact. If (u_k) converges weakly as $k\to+\infty$, then it converges strongly and the limit belongs to $\mathcal{S}$.*

Proof. If (u_k) converges weakly as $k\to+\infty$, then it is bounded. By Proposition 6, the weak limit belongs to $\mathcal{S}\neq\emptyset$ and $\lim_{k\to+\infty}\Theta(u_k)=\min\Theta$. Therefore, there exist $r\in\mathbf{R}$ and $R>0$ such that the set Γ_r^R contains (u_k) and is relatively compact for the strong topology. A standard argument shows that (u_k) converges strongly as $k\to+\infty$. ■

3.2. **Convergence of the sequence** (y_k)**.** Let us turn our attention to the auxiliary sequence (y_k), whose convergence will result from a detailed analysis of the second equation in (IFB):

$$\frac{y_{k+1}-y_k}{\lambda}+\nabla\Psi(u_{k+1})-au_{k+1}+by_{k+1}=0. \tag{34}$$

We have the following:

Proposition 12. *Let* (u_k,y_k) *satisfy* (IFB)*. The following holds:*

i) *If* (u_k) *converges weakly to some* u_∞ *and* $\nabla\Psi$ *is* weak-to-weak *sequentially continuous, then* (y_k) *converges weakly to some* y_∞*;*

ii) *If* (u_k) *converges strongly to some* u_∞*, then* (y_k) *converges strongly to some* y_∞*.*

In either case, we have

$$\nabla\Psi(u_\infty)-au_\infty+by_\infty=0. \tag{35}$$

We shall use the following result concerning real sequences:

Lemma 13. *Let* (p_k) *and* (ε_k) *be real sequences such that* $\limsup_{k\to+\infty}\varepsilon_k\leq 0$ *and*

$$p_{k+1}\leq \eta p_k+\varepsilon_k$$

for some $\eta\in(0,1)$ *and all sufficiently large* k*. Then* $\limsup_{k\to+\infty}p_k\leq 0$*.*

Proof. A simple induction argument shows that

$$p_k\leq \eta^k p_0+\sum_{j=0}^{k-1}\eta^{k-j-1}\varepsilon_j.$$

Set $C_k=\sum_{j=0}^{k-1}\varepsilon_j\eta^{k-j-1}$. Since $\lim_{k\to+\infty}\eta^k p_0=0$, it suffices to prove that $\limsup_{k\to+\infty}C_k\leq 0$. To this end, fix $\varepsilon>0$. Since $\limsup_{k\to+\infty}\varepsilon_k\leq 0$, there exists $k_\varepsilon>0$ such that $\varepsilon_j\leq\varepsilon$ for all $j\geq k_\varepsilon$. Then, for all $k>k_\varepsilon$ we have

$$C_k\leq\sum_{j=0}^{k_\varepsilon-1}\eta^{k-j-1}\varepsilon_j+\sum_{j=k_\varepsilon}^{k}\eta^{k-j-1}\varepsilon_j\leq\eta^{k-1}M_\varepsilon+\frac{\varepsilon}{1-\eta},$$

where the quantity $M_\varepsilon=\sum_{j=0}^{k_\varepsilon-1}\eta^{-j}\varepsilon_j$ is independent of k. Letting $k\to+\infty$ we deduce that

$$\limsup_{k\to+\infty}C_k\leq\frac{\varepsilon}{1-\eta}.$$

This being true for any $\varepsilon>0$, the conclusion follows. ■

Proof of Proposition 12. For $k\geq 1$, set

$$bw_k=au_{k+1}-\nabla\Psi(u_{k+1}). \tag{36}$$

Let u_∞ be the (weak or strong) limit of the sequence (u_k) as $k\to+\infty$ and set

$$by_\infty=au_\infty-\nabla\Psi(u_\infty),$$

so that (34) can be rewritten as

$$(y_{k+1}-y_\infty)=\eta(y_k-y_\infty)+\lambda b\eta(w_k-y_\infty), \tag{37}$$

where $\eta=\frac{1}{1+\lambda b}\in(0,1)$.

In order to prove *i*), let (u_k) converge weakly to u_∞ as $k\to+\infty$. Since $\nabla\Psi$ is *weak-to-weak* sequentially continuous, the sequence (w_k) converges weakly to y_∞ as $k\to+\infty$. Take $\xi\in H$ and multiply equation (37) by ξ to obtain

$$\langle y_{k+1}-y_\infty,\xi\rangle=\eta\langle y_k-y_\infty,\xi\rangle+\lambda b\eta\langle w_k-y_\infty,\xi\rangle.$$

Use Lemma 13 with $p_k = \langle y_k - y_\infty, \xi\rangle$ and $\varepsilon_k = \lambda b\eta\langle w_k - y_\infty, \xi\rangle$ to deduce that

$$\limsup_{k\to+\infty}\langle y_k - y_\infty, \xi\rangle \le 0$$

for all $\xi \in H$. Replacing ξ by $-\xi$ we conclude that

$$\liminf_{k\to+\infty}\langle y_k - y_\infty, \xi\rangle \ge 0$$

and so (y_k) converges weakly to y_∞.

To prove $ii)$, use (37) and the triangle inequality to deduce that

$$\|y_{k+1} - y_\infty\| \le \eta\|y_k - y_\infty\| + \lambda b\eta\|w_k - y_\infty\|.$$

Recall that $\lim_{k\to+\infty}\|w_k - y_\infty\| = 0$ if (u_k) converges strongly to u_∞ as $k \to +\infty$, and use Lemma 13 with $p_k = \|y_k - y_\infty\|$ and $\varepsilon_k = \lambda b\eta\|w_k - y_\infty\|$ to conclude that $\lim_{k\to+\infty}\|y_k - y_\infty\| = 0$. ■

4. Applications

One of the main advantages of forward-backward algorithms is that the backward (implicit) step can handle nondifferentiable functions, such as those commonly encountered in hard-constrained problems, or sparse approximation and compressed sensing. Let us examine some problems of this kind.

4.1. **The case $\Phi = \delta_C$.** When $\Phi = \delta_C$ is the indicator function of a nonempty closed convex set C, the proximity operator of Φ is the projection operator onto C, Proj_C. For simplicity, we take $a = b = 1$. Algorithm (IFB), as formulated in (6) becomes

$$\left\{\begin{array}{rcl} u_{k+1} &=& \mathrm{Proj}_C\left(\lambda\, y_k + (1-\lambda)\, u_k\right) \\ y_{k+1} &=& (1-\mu)\, y_k + \mu\left(u_{k+1} - \nabla\Psi(u_{k+1})\right), \end{array}\right. \tag{38}$$

where $\mu = \frac{\lambda}{1+\lambda}$, and will be referred to as the gradient-projection algorithm. Figure 1 illustrates how the two subiterations are performed. The thick broken line represents part of the boundary of the set C, while the ellipses depict the level sets of the function Ψ.

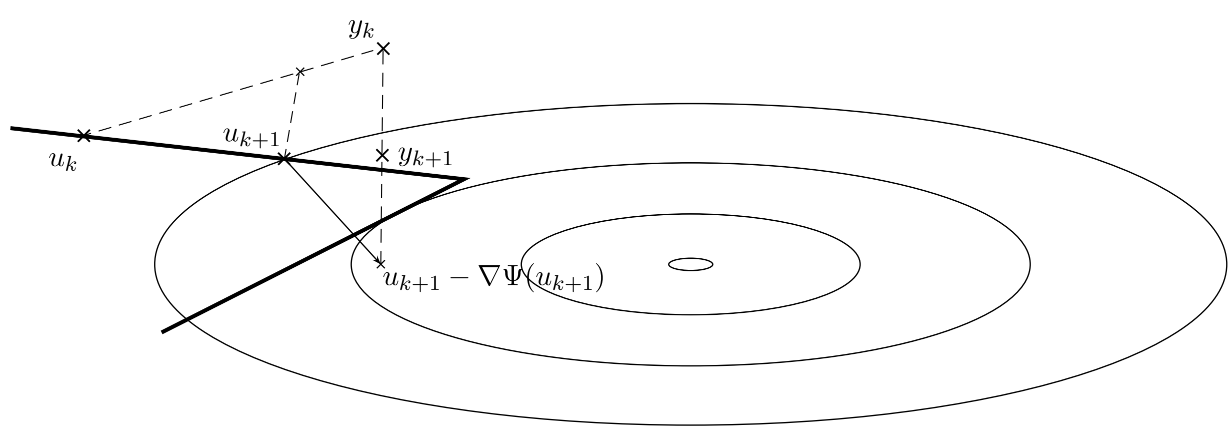

Figure 1

Theorem 1 and Proposition 12 guarantee convergence of the sequence (u_k, y_k) provided the stepsize λ satisfies $0 < \lambda < \Lambda = \min\left\{1, \frac{4}{L_\Psi}\right\}$.

4.1.1. *Hard-constrained feasibility problems.* The convex feasibility problem is to find a point in the intersection of a finite collection of nonempty closed convex subsets $C, C_1, \dots, C_p$ of a real Hilbert space H; see [13, 26]. In many instances, this intersection may turn out to be empty, and a relaxation of this problem in the presence of a hard constraint represented by C is to

$$\min \left\{ \frac{1}{2} \sum_{i=1}^{p} \omega_i \text{dist}(v, C_i)^2 : \ v \in C \right\} \tag{39}$$

where ω_i are positive constants such that $\sum_i \omega_i = 1$; see [27]. For example, hard constraints can model the positivity of some physical or economical variables, while soft constraints (modeled by the C_i) can reflect some noisy data, measurements. By applying a forward-backward splitting algorithm to this problem, we aim at finding a point which satisfies the hard constraint C, while the other constraints are satisfied in a possibly weaker sense (see [29] and references therein). Set

$$\Psi(v) = \frac{1}{2} \sum_{i=1}^{p} \omega_i \text{dist}(v, C_i)^2. \tag{40}$$

By a standard convex analysis result, each function $\Psi_i(v) = \frac{1}{2}\text{dist}(v, C_i)^2$ is C^1 convex, and its gradient, equal to $\nabla \Psi_i(v) = v - \text{Proj}_{C_i}(v)$, is Lipschitz continuous with Lipschitz constant equal to 1. By convex combination, the same property holds true for Ψ, and we can take $L_\Psi = 1$ as a Lipschitz constant of $\nabla \Psi$. Take $\Phi = \delta_C$, the indicator function of C.
The classical forward-backward splitting algorithm (gradient-projection) reads:
Take $u_0 \in \mathbf{R}^n$ and $0 < \lambda < 2$. For $k = 0, 1, \dots$,

$$u_{k+1} = \text{Proj}_C \left((1-\lambda) u_k + \lambda \sum_{i=1}^{p} \omega_i \text{Proj}_{C_i}(u_k) \right). \tag{41}$$

Under the assumption that the solution set of problem (39) is nonempty, (u_k) converges weakly to an optimal solution of (39); see [30, Corollary 4.10].
Let us now analyze the (IFB) gradient-projection algorithm. We successively analyze different choices for the parameters a and b (from a simple case to the general case):

i) Let us first take $a = b = 1$. Set $\mu = \frac{\lambda}{1+\lambda}$. Taking account of the definition of Ψ, Algorithm (38) gives

$$\begin{cases} u_{k+1} &= \text{Proj}_C \left((1-\lambda) u_k + \lambda y_k \right) \\ y_{k+1} &= (1-\mu)\, y_k + \mu \sum_{i=1}^{p} \omega_i \text{Proj}_{C_i}(u_{k+1}). \end{cases}$$

One can observe that in the classical gradient-projection algorithm we have

$$u_{k+1} = \text{Proj}_C \left((1-\lambda) u_k + \lambda y_k \right) \tag{42}$$

with $y_k = \sum_{i=1}^{p} \omega_i \text{Proj}_{C_i}(u_k)$, while, in the (IFB) algorithm, relation (42) holds with $y_k = (1-\mu)\, y_{k-1} + \mu \sum_{i=1}^{p} \omega_i \text{Proj}_{C_i}(u_k)$, hence the inertial features.
Since $L_\Psi = 1$, by application of Theorem 1, with $0 < \lambda < 1$, and assuming that the solution set of problem (39) is non empty, then (u_k) converges weakly to an optimal solution of (39). Note that with $L_\Psi = 1$, the choice $a = b = 1$ does not improve the step size limitation on λ with respect to the classical forward-backward splitting algorithm.

ii) Let us now take $a = b$ and play on the common value of the two parameters. The (IFB) algorithm reads:

$$\begin{cases} u_{k+1} &= \text{Proj}_C \left((1-\lambda a) u_k + \lambda a y_k \right) \\ y_{k+1} &= \frac{1}{1+\lambda a} y_k + \frac{\lambda(a-1)}{1+\lambda a} u_{k+1} + \frac{\lambda}{1+\lambda a} \sum_{i=1}^{p} \omega_i \text{Proj}_{C_i}(u_{k+1}). \end{cases}$$

By taking $a = \frac{1}{4}$, convergence of the above algorithm holds under the condition $0 < \lambda < 4$. One obtains $0 < \frac{\lambda}{1+\lambda a} < 2$, hence a limitation on the gradient term coefficient which is similar to the classical forward-backward splitting algorithm.

iii) Take now general positive parameters a and b. The (IFB) algorithm reads:

$$\left\{\begin{array}{lcl} u_{k+1} & = & \text{Proj}_C\left((1-\lambda a)u_k+\lambda b y_k\right) \\ y_{k+1} & = & \frac{1}{1+\lambda b}y_k+\frac{\lambda a}{1+\lambda b}u_{k+1}-\frac{\lambda}{1+\lambda b}\left(u_{k+1}-\sum_{i=1}^p\omega_i\text{Proj}_{C_i}(u_{k+1})\right). \end{array}\right. \tag{43}$$

Whenever $0<a<\frac{1}{2}$, by choosing $b\le\frac{2a^2}{1-2a}$, we obtain $\Lambda=\frac{1}{a}$. Hence, Λ can be made arbitrarily large (consequently, λ can be taken arbitrarily large) by choosing a (reasonably) small. So doing one can take the gradient term coefficient, $\frac{\lambda}{1+\lambda b}$, arbitrarily large.

Let us illustrate this fact by taking a (reasonably) small, $b=2a^2$, and $\lambda=\frac{1}{2a}$. Hence $\lambda a=\frac{1}{2}$, $\lambda b=a$ and the algorithm becomes

$$u_{k+1}=\text{Proj}_C\left(\frac{1+2a}{2(1+a)}u_k+\frac{a}{1+a}u_{k-1}-\frac{1}{2(1+a)}\left(u_k-\sum_{i=1}^p\omega_i\text{Proj}_{C_i}(u_k)\right)\right). \tag{44}$$

Thus, when a is small there is a compensation between the large coefficient λ and the correspondingly small coefficients (a and b). The algorithm in that case is close to the classical algorithm with $\lambda=\frac{1}{2}$, plus the inertial effect. So doing, when L_Ψ is moderate (here it is equal to 1), the (IFB) algorithm provides a new way to relax the forward-backward algorithm (compare with [29]), and thus gives some further flexibility to this class of algorithms.

In the next examples we analyze situations where L_Ψ may be large.

4.1.2. *An inertial CQ algorithm.* The algorithms of the previous section, e.g. (41, 44), are implementable only if the projection operators Proj_C, $(\text{Proj}_{C_i})_{i\in\{1,\dots,p\}}$ are available. But in several instances, computing the projection onto a convex set may be quite an issue. As a typical situation, consider the so-called *split feasibility problem* (see e.g. [24]): let $A:H\to Y$ be a linear continuous operator from H into another Hilbert space Y and let $Q\subset Y$ be a closed convex set. Given a hard (closed convex) constraint $C\subseteq H$, the split feasibility problem is to

$$\text{find}\quad u\in C\quad\text{such that}\quad Au\in Q\ , \tag{45}$$

and it is assumed that Proj_C and Proj_Q are easily calculated.

As a common example, take for Q the closed ball of center $b\in Y$ and radius ε. The problem is then to find $u\in C$ satisfying the linear constraint $Au=b$ up to some tolerance, i.e. $\|Au-b\|\le\varepsilon$.

With C_1 denoting the convex closed set $\{v\in H: Av\in Q\}$, problem (45) is obviously equivalent to

$$\text{find}\quad u\in C\cap C_1\ .$$

So the split feasibility problem is a particular case of the hard-constrained feasibility problem. But using the function $v\mapsto\frac{1}{2}\text{dist}(v,C_1)^2$ as a penalty to relax the problem turns out to be awkward, since it involves the projection operator Proj_{C_1} which may be hard to compute (and would likely require some sort of inverse for the operator A^tA, with $A^t:Y\to H$ the transpose of A). Instead, it is more advantageous to use a penalty function which exploits the particular structure of the set C_1, namely

$$\Psi(v)=\frac{1}{2}\text{dist}(Av,Q)^2. \tag{46}$$

Clearly, Ψ is a convex differentiable function whose gradient is equal to

$$\nabla\Psi=A^t(I-\text{Proj}_Q)A, \tag{47}$$

The operator $\nabla\Psi:H\to H$ is Lipschitz continuous with Lipschitz constant

$$L_\Psi\le\rho(A^tA), \tag{48}$$

where $\rho(A^tA)$ is the spectral radius of A^tA; see [22]. Recall that, in the finite dimensional case, $\rho(A^tA)$ is the largest eigenvalue of the nonnegative symmetric matrix A^tA.

Given the hard constraint $C \subset H$, the split feasibility problem (45) can be relaxed to

$$\min \left\{ \delta_C(v) + \frac{1}{2}\mathrm{dist}(Av, Q)^2 : \ v \in H \right\}. \tag{49}$$

The classical forward-backward algorithm, when applied to this structured minimization problem gives rise to the so-called CQ algorithm, originally introduced by Byrne in [22]:

$$u_{k+1} = \mathrm{Proj}_C \left(u_k - \lambda A^t (I - \mathrm{Proj}_Q) A u_k \right). \tag{50}$$

Assuming that the solution set of the minimization problem (49) is nonempty, and $\lambda < 2/\rho(A^t A)$, this algorithm generates sequences (u_k) weakly converging to solutions of (49); see [30, Corollary 4.10], and [21, 22] for the finite-dimensional case.

Let us now consider the (IFB) algorithm and apply it to the structured minimization problem (49):

$$\left\{ \begin{array}{rcl} u_{k+1} & = & \mathrm{Proj}_C \left((1 - \lambda a) u_k + \lambda b y_k \right) \\ y_{k+1} & = & \frac{1}{1+\lambda b} y_k + \frac{\lambda a}{1+\lambda b} u_{k+1} - \frac{\lambda}{1+\lambda b} \left(u_{k+1} - A^t (I - \mathrm{Proj}_Q) A(u_{k+1}) \right). \end{array} \right. \tag{51}$$

Assuming that the solution set of the minimization problem (49) is nonempty, and

$$0 < \lambda < \Lambda = \min \left\{ \frac{1}{a}, \frac{2(a+b)}{b\rho(A^t A)} \right\}$$

with $a, b > 0$, this algorithm generates sequences (u_k) weakly converging to solutions of (49).

4.2. **Pareto front in an infinite-dimensional nonconvex setting.** According to the Principle of Least Action, unperturbed mechanical systems tend to stabilize at minimizers of a given energy function Φ (the *action*). In many typical yet relevant model examples this energy function is a (symmetric and) positive-definite quadratic form in some L^2 space, thus having a unique minimizer. When selecting a configuration of the system such that a given (not necessarily convex) cost Ψ is reasonably small, while respecting − to some extent − the natural equilibrium of the system, a decision maker is led to a multiobjective optimization problem. A sensible notion of solution for this problem is a *Pareto equilibrium* − where neither the action nor the cost can be reduced without increasing the value of the other function.

Let A, B be bounded, self-adjoint and linear operators on a Hilbert space H. Assume A is uniformly elliptic; that is, there exists an *ellipticity coefficient* $e > 0$ such that

$$\langle Ax, x \rangle \geq e\|x\|^2$$

for all $x \in H$ (the smallest eigenvalue in the case of a positive-definite symmetric matrix). Let $a, b \in H$ and $\alpha, \beta \in \mathbf{R}$ and let C be a nonempty, closed and convex subset of H. The function $\Phi : H \to \mathbf{R} \cup \{+\infty\}$ defined by

$$\Phi(x) = \frac{1}{2}\langle Ax, x \rangle + \langle a, x \rangle + \alpha + \delta_C(x) \qquad \text{for } x \in H,$$

is proper, lower-semicontinuous and strongly convex. In turn, the function $\Psi : H \to \mathbf{R}$ defined by

$$\Psi(x) = \frac{1}{2}\langle Bx, x \rangle + \langle b, x \rangle + \beta \qquad \text{for } x \in H,$$

is continuously differentiable. Its gradient $\nabla\Psi$, given by $\nabla\Psi(x) = Bx + b$, is Lipschitz-continuous with constant $\|B\|$ (the operator norm of B) and, being linear, it is also weak-to-weak continuous. Let $\eta \in (0, 1)$ and define $\Omega_\eta : H \to \mathbf{R} \cup \{+\infty\}$ by

$$\Omega_\eta(x) = \eta\, \Phi(x) + (1 - \eta)\, \Psi(x) \qquad \text{for } x \in H.$$

Every minimizer of Ω_η on C is a Pareto equilibrium for the multiobjective problem posed in C.

Remark 14. It is to be noted that, in general, some parts of the Pareto front may not be recoverable by this procedure if Ψ is not convex (see [32]).

Set $\eta_0 = \|B\|e^{-1}$. It is easy to see that if $\eta > \frac{\eta_0}{\eta_0+1}$, the function Ω_η is convex, bounded from below and has minimizers. The preceding discussion allows us to use part iii) of Theorem 1 to guarantee the convergence of (IFB).

Remark 15. If Ψ is convex (not necessarily quadratic but verifying Hypothesis $\mathbf{H}_\Psi$), one may use part iv) of Theorem 1.

4.3. **Signal recovery via l^1 regularization.** Forward-backward optimization algorithms have been successfully applied to a wide class of signal and image processing problems ranging from restoration and reconstruction to synthesis and design, see [29, 16, 22] and references therein.
A basic linear inverse problem is to estimate an unknown signal u satisfying the relation

$$Au = b + w,$$

where $A \neq 0$ is an $m \times n$ real matrix and $b \in \mathbf{R}^m$. Both A and b are known, while w is an unknown noise vector. In order to recover the signal u from the noisy measurement b, a common approach for this estimation problem is to solve the regularized least squares minimization problem

$$(RLS) \quad \min\{\gamma\mathcal{R}(v) + \|Av - b\|^2 : \quad v \in \mathbf{R}^n\},$$

where $\mathcal{R}$ is an operator which regularizes, stabilizes the problem, and γ is a positive parameter. In compressive sensing, in order to recover sparse solutions of underdetermined linear systems, one usually takes $\mathcal{R}$ equal to the l^1 norm of $\mathbf{R}^n$ (see [31], [23]), and thus one considers the l^1 regularization of the inverse problem:

$$\min\{\gamma\|v\|_1 + \|Av - b\|^2 : \quad v \in \mathbf{R}^n\}. \tag{52}$$

This is a structured minimization problem with $\Phi(v) = \gamma\|v\|_1$ a convex nondifferentiable function, and $\Psi(v) = \|Av - b\|^2$ a convex C^1 function. The crucial point is that the proximity mapping of the l^1 norm can be computed explicitly, via elementary operations. It is just the shrinkage or soft threshold operator defined by

$$\left(\mathrm{prox}_{\lambda\gamma\|\cdot\|_1}(v)\right)_i = (|v_i| - \lambda\gamma)^+ \mathrm{sgn}(v_i)\ ,\ i = 1, \dots, n. \tag{53}$$

for some positive parameter λ. Equivalently,

$$\left(\mathrm{prox}_{\lambda\gamma\|\cdot\|_1}(v)\right)_i = \begin{cases} v_i - \lambda\gamma & \text{if} \quad v_i > \lambda\gamma \\ 0 & \text{if} \quad |v_i| \leq \lambda\gamma \\ v_i + \lambda\gamma & \text{if} \quad v_i < -\lambda\gamma. \end{cases} \tag{54}$$

The classical forward-backward operator provides the so-called iterative shrinkage/thresholding algorithm (ISTA) (see [15, 16])

$$u_{k+1} = \mathrm{prox}_{\lambda\gamma\|\cdot\|_1}\left(u_k - 2\lambda A^t(Au_k - b)\right). \tag{55}$$

Let us now consider the (IFB) algorithm and apply it to the structured minimization problem (52):

$$\begin{cases} u_{k+1} &= \mathrm{prox}_{\lambda\gamma\|\cdot\|_1}\left((1-\lambda a)u_k + \lambda b y_k\right) \\ y_{k+1} &= \frac{1}{1+\lambda b}y_k + \frac{\lambda a}{1+\lambda b}u_{k+1} - \frac{2\lambda}{1+\lambda b}A^t(Au_{k+1} - b). \end{cases} \tag{56}$$

Noticing that the solution set of the minimization problem (52) is nonempty, and taking

$$0 < \lambda < \Lambda = \min\left\{\frac{1}{a}, \frac{2(a+b)}{bL_\Psi}\right\}$$

with $L_\Psi = 2\rho(A^tA)$ and $a, b > 0$, this algorithm generates sequences (u_k) converging to solutions of (52).

A numerical illustration. We compared the IFB algorithm (56) with ISTA (55) and FISTA algorithms (see [15]). We also introduced an accelerated version of IFB, called FIFB, patterned after [35, 15] (that is (u_{k+1}, y_{k+1}) is obtained by applying IFB to $(u_k, y_k) + \frac{t_k-1}{t_{k+1}}(u_k - u_{k-1}, y_k - y_{k-1})$ where (t_k) is Nesterov's sequence). Although convergence of FIFB does not follow from the present analysis, it seems fair to compare ISTA to IFB and FISTA to an accelerated, though empirical, version of IFB.

Numerical computations were performed on the well-known 256x256 cameraman test image in the same conditions as in [15]: the observed image is derived from the original one through a Gaussian 9x9 blur with standard deviation 4 followed by an additive zero-mean Gaussian noise with standard deviation 10^{-3} (see fig. 2).

Figure 2. Original (left) and observed (right) image.

The corruption operator is $A = RW$ where R is the blur operator with reflexive boundary conditions and W is the inverse of the three-stage Haar wavelet transform. We note $L_\Psi = 2\rho(A^t A) = 2$ and $\gamma = 2.10^{-5}$. The initial image u_0 was the observed image and, where relevant, y_0 was zero.

In the first trials, we noted that for reasonable values of parameters (a, b, λ), IFB outperformed ISTA but was outperformed by FISTA. In turn, FIFB was comparable to, but slightly slower than FISTA. However, by trial and error, we were able to obtain some particular values of the parameters for which FIFB was faster than FISTA. Accelerating IFB certainly deserves further study, but this goes beyond the scope of the present paper.

Fig. 3 compares the function values $\Phi(u_k) + \Psi(u_k)$ given by the various methods for the first 200 iterations. Algorithms IFB and FIFB were applied to a dozen of values of the parameters and the best cases were retained: $(a, b, \lambda) = (L_\Psi/4, L_\Psi/4, 3.9/L_\Psi)$ for IFB and $(a, b, \lambda) = (L_\Psi/4, 4L_\Psi, 1.5/L_\Psi)$ for FIFB.

Fig. 4 compares the restored images after 200 iterations of the various algorithms. Differences between ISTA and IFB on the one hand and FISTA and FIFB on the other hand are hardly noticeable. It is important to mention, though, that FISTA guarantees the convergence of the function values $(\Phi(u_k) + \Psi(u_k))$ but the convergence of (u_k) is still an open question. For IFB, both the convergence of the function values and the iterates are obtained.

5. Conclusion

In this paper we considered the problem of minimizing a convex function Θ, that can be decomposed as the sum of a convex function Φ and a smooth function Ψ. To this end, we introduced a class of forward-backward algorithms with inertial features. These algorithms can be interpreted as discretizations of

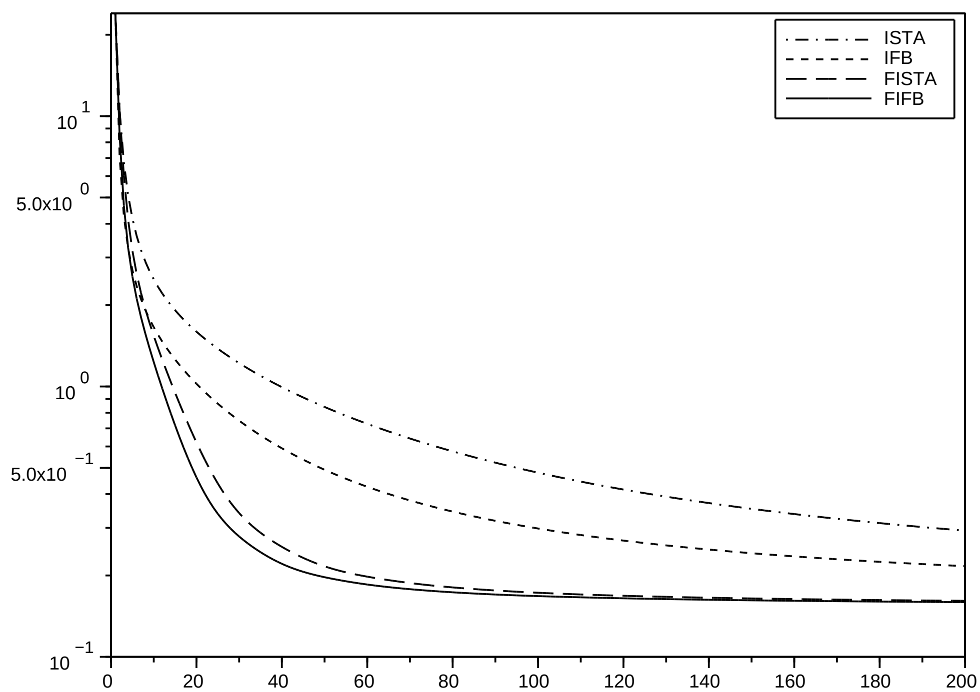

FIGURE 3. Function values $\Phi(u_k) + \Psi(u_k)$

the second-order dynamical system with Hessian-driven damping studied in [11]. Convergence can be guaranteed in a variety of cases that go beyond the standard $2/L_\Psi$ bound for the step sizes in the classical forward-backward method, and also account for a possibly nonconvex function Ψ.
Further research may concern:

(1) The theoretical and computational effect of acceleration schemes, possibly in the line of [35], [15] and [16];
(2) The convergence analysis in a fully nonconconvex and nonsmooth framework using, for instance, the techniques in [7];
(3) Extensive numerical tests, not only to assess the computational power of the method and compare it with benchmark algorithms, but also to obtain an insight into the optimal selection of the parameters.

This prospective work goes beyond the scope of this paper and therefore opens potential directions for future studies.

References

[1] F. Alvarez, *On the minizing property of a second order dissipative dynamical system in Hilbert spaces*, SIAM J. Control Optim., 39 (2000), pp. 1102-1119.
[2] F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM J. Optim., 3 (2004), pp. 773-782.
[3] F. Alvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Var. Anal., 9 (2001), pp. 3-11.
[4] F. Alvarez, H. Attouch, J. Bolte, and P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl., 81 (2002), pp. 747–779.

FIGURE 4. Restored images; clockwise from top left: ISTA, FISTA, FIFB, IFB

[5] F. ALVAREZ, J.M. PEREZ, *Asymptotic analysis of evolution equations associated with Newton's method for parametric approximations of convex minimization problems*, Appl. Math. Optim. 38 (1998), 193-217.

[6] A. S. ANTIPIN, *Minimization of convex functions on convex sets by means of differential equations*, Differ. Equ., 30 (1994), pp. 1365–1375.

[7] 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*, Math. Program. 137 (2013), pp. 91–129.

[8] H. ATTOUCH, L. M. BRICEÑO-ARIAS, AND P. L. COMBETTES, *A parallel splitting method for coupled monotone inclusions*, SIAM J. Control Optim., 48 (2010), pp. 3246–3270.

[9] H. ATTOUCH, A. CABOT, P. REDONT, *The dynamics of elastic shocks via epigraphical regularization of a differential inclusion*, Adv. Math. Sci. Appl., 12 (2002), pp. 273–306.

[10] H. ATTOUCH, M.-O. CZARNECKI, AND J. PEYPOUQUET, *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM J. Optim., 21 (2011), pp. 1251–1274.

[11] H. ATTOUCH, P.-E. MAINGÉ AND P. REDONT, *A second-order differential system with Hessian-driven damping; application to non-elastic shock laws*, Differ. Equ. Appl., 4 (2012), pp. 27–65.

[12] J.-P. AUBIN, A. CELLINA, *Differential Inclusions*, Springer, 1984.

[13] H. H. BAUSCHKE AND J. M. BORWEIN, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.

[14] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

[15] A. Beck and M. Teboulle, *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems*, SIAM J. on Imaging Sci., 2 (2009), pp. 183–202.
[16] A. Beck and M. Teboulle, *Gradient-based algorithms with applications in signal recovery problems*, In Convex Optimization in Signal Processing and Communications, D. Palomar and Y. Eldar Eds., pp. 33–88. Cambridge University Press, 2010.
[17] D. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont MA, 1999.
[18] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland/Elsevier, New-York, 1973.
[19] L. M. Briceño-Arias and P. L. Combettes, *A monotone+skew splitting model for composite monotone inclusions in duality*, SIAM J.Optim., 21 (2011), pp. 1230–1250.
[20] R.E. Bruck, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces*, J. Funct. Anal., 18 (1975), pp. 15–26.
[21] C. Byrne, *Iterative oblique projection onto convex sets and the split feasibility problem*, Inverse Problems, 18 (2002), pp. 441–453,
[22] C. Byrne, *A unified treatment of some iterative algorithms in signal processing and image reconstruction*, Inverse Problems, 20 (2004), pp. 103–120.
[23] E. J. Candès, J. K. Romberg and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.
[24] Y. Censor and T. Elfving, *A multiprojection algorithm using Bregman projections in a product space*, Numer. Algorithms, 8 (1994) pp. 221–239.
[25] G. Chen and M. Teboulle, *A proximal-based decomposition method for convex minimization problems*, Math. Program., 64 (1994), pp. 81–101.
[26] P. L. Combettes, *Hilbertian convex feasibility problem: Convergence of projection methods*, Appl. Math. Optim., 35 (1997), pp. 311–330.
[27] P. L. Combettes and P. Bondon, *Hard-constrained inconsistent signal feasibility problems*, IEEE Trans. Signal Process., 47 (1999), pp. 2460–2468.
[28] P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization, 53 (2004), pp. 475–504.
[29] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing, in: Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, (H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Editors), pp. 185-212. Springer, New York, 2011.
[30] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation, 4 (2005), pp. 1168–1200.
[31] D. L. Donoho, *Compressed Sensing*, IEEE Trans. Inform. Theory, 4 (2006), 1289–1306.
[32] A. M. Geoffrion, *Proper efficiency and the theory of vector maximization*, J. Math. Anal. Appl., 22 (1968), 618–630.
[33] P.L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.
[34] Y. E. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer, Boston, 2004.
[35] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.
[36] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics Appl. Math. 30, SIAM, Philadelphia, 2000.
[37] G.B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Anal. Appl., 72 (1979), pp. 383–390.
[38] J. Peypouquet and S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, J. of Convex Analysis, 17 (2010), pp. 1113–1163.
[39] B. Polyak, *Introduction to Optimization*, Optimization Software, New-York, 1987.
[40] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 15 (1976), pp. 877-898.
[41] M. V. Solodov and B. F. Svaiter, *A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator*, Set-Valued Var. Anal., 7 (1999), pp. 323–345.
[42] P. Tseng, *Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming*, Math. Program., 48 (1990), pp. 249–263.
[43] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138.
[44] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446.

Institut de Mathématiques et Modélisation de Montpellier, UMR 5149 CNRS, Université Montpellier 2, place Eugène Bataillon, 34095 Montpellier cedex 5, France
E-mail address: attouch@math.univ-montp2.fr

Departamento de Matemática, Universidad Técnica Federico Santa María, Avenida España 1680, Valparaíso, Chile.
E-mail address: juan.peypouquet@usm.cl

Institut de Mathématiques et Modélisation de Montpellier, UMR 5149 CNRS, Université Montpellier 2, place Eugène Bataillon, 34095 Montpellier cedex 5, France
E-mail address: redont@math.univ-montp2.fr