

# An inertial proximal-gradient penalization scheme for constrained convex optimization problems

Radu Ioan Boţ · Ernő Robert Csetnek · Nimit Nimana

Received: date / Accepted: date

**Abstract** We propose a proximal-gradient algorithm with penalization terms and inertial and memory effects for minimizing the sum of a proper, convex and lower semicontinuous and a convex differentiable function subject to the set of minimizers of another convex differentiable function. We show that, under suitable choices for the step sizes and the penalization parameters, the generated iterates weakly converge to an optimal solution of the addressed bilevel optimization problem, while the objective function values converge to its optimal objective value.

**Keywords** proximal-gradient algorithm · inertial algorithm · penalization · Fenchel conjugate

**Mathematics Subject Classification (2000)** 47H05 · 65K05 · 90C25

The work of the first author was partially supported by by FWF (Austrian Science Fund), project I 2419-N32. The work of the second author was supported by by FWF (Austrian Science Fund), project P 29809-N32. Research done during the two months' stay of the third author in Spring 2016 at the Faculty of Mathematics of the University of Vienna. The third author is thankful to the Royal Golden Jubilee PhD Program for financial support.

Radu Ioan Boţ  
University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria  
Tel.: +43 1 4277 50772  
Fax: +43 1 4277 850772  
E-mail: radu.bot@univie.ac.at

Ernö Robert Csetnek  
University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria  
Tel.: +43 1 4277 50768  
Fax: +43 1 4277 850768  
E-mail: ernoe.robert.csetnek@univie.ac.at

Nimit Nimana  
Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok, 65000, Thailand  
E-mail: nimitn@hotmail.com

## 1 Introduction and preliminaries

Let  $\mathcal{H}$  be a real Hilbert space with inner product  $\langle \cdot, \cdot \rangle$  and associated norm  $\|\cdot\|$ . Let  $f : \mathcal{H} \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$  be a proper, convex and lower semicontinuous function, and  $h : \mathcal{H} \rightarrow \mathbb{R}$  and  $g : \mathcal{H} \rightarrow \mathbb{R}$  be convex and differentiable functions with Lipschitz continuous gradients with positive Lipschitz constants  $L_h$  and  $L_g$ , respectively. We consider the bilevel optimization problem

$$\min_{x \in \arg \min g} f(x) + h(x) \quad (1)$$

and assume that the set

$$\mathcal{S} := \arg \min \{f(x) + h(x) : x \in \arg \min g\}$$

is nonempty. We also assume without loss of generality that  $\min g = 0$ .

The work of Attouch and Czarnecki [5] has represented the starting point of a series of articles approaching the minimization of a smooth or (sometimes) a complexly structured nonsmooth objective function subject to the set of minimizers of another function, from either a discrete perspective through iterative numerical algorithms or a continuous one through dynamical systems (see [4–8, 11, 17–19, 24–27, 30, 37, 39]). The function determining the feasible set is evaluated in both settings in the spirit of penalty methods and contributes to the convergence of the generated sequences, in the discrete setting, and to the asymptotic convergence of the generated trajectories, in the continuous setting, to an optimal solution of the underlying bilevel optimization problem. We emphasize in particular the proximal-gradient algorithm with penalty term which has been introduced in [8] and for which weak ergodic convergence has been proved.

In this paper we consider this algorithm in the context of solving problem (1) and we enhance it with inertial and memory effects. Our aim is to provide suitable choices for the step sizes and the penalization parameters, such that the generated iterates weakly converge to an optimal solution of the (1), while the objective function values converge to its optimal objective value. Algorithms of inertial type follow by time discretization of differential inclusions of second order type (see [1, 3]) and have been first investigated in the context of the minimization of a differentiable function by Polyak in [40] and Bertsekas in [14]. In the last two decades intensive research efforts dedicated to algorithms of inertial type and to their convergence behaviour can be noticed (see [1–3, 10, 20–23, 28, 29, 31–36, 38]). For a variety of situations, in particular in the context of solving real-world problems, the presence of inertial terms improves the convergence behavior of the generated sequences. It is also well-known (see [9, 13]) that enhancing the proximal-gradient algorithm with inertial effects may lead to a considerable improvement of the convergence behaviour of the sequence of objective function values.

The proximal-gradient algorithm with penalization terms and inertial and memory effects we propose for solving (1) follows.

**Algorithm 1** *Initialization:* Choose the positive sequences  $\{\lambda_n\}_{n=1}^\infty$ ,  $\{\beta_n\}_{n=1}^\infty$ , and a nonnegative constant  $\alpha \in [0, 1)$ . Take arbitrary  $x_0, x_1 \in \mathcal{H}$ .

*Iterative step:* For every  $n \geq 1$  and given current iterates  $x_{n-1}, x_n \in \mathcal{H}$  define  $x_{n+1} \in \mathcal{H}$  by

$$x_{n+1} := \text{prox}_{\lambda_n f}(x_n + \alpha(x_n - x_{n-1}) - \lambda_n \nabla h(x_n) - \lambda_n \beta_n \nabla g(x_n)).$$

For  $x \in \mathcal{H}$  we denote by  $\text{prox}_{\lambda_n f}(x)$  the proximal point of the function  $f$  of parameter  $\lambda_n$  at  $x$ , which is the unique optimal solution of the optimization problem

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

In Algorithm 1,  $\{\lambda_n\}_{n=1}^\infty$  denotes the sequence of step sizes,  $\{\beta_n\}_{n=1}^\infty$  the sequence of penalization parameters, and  $\alpha \in [0, 1)$  the parameter that controls the inertial terms.

The proposed numerical scheme recovers, when  $\alpha = 0$ , the algorithm investigated in [37] and, under the additional assumption  $f = 0$ , the gradient method of penalty type from [39]. In case  $f = 0$ , Algorithm 1 gives rise to the gradient method of penalty type with inertial and memory effects introduced and studied in [30].

We prove weak convergence for the generated iterates to an optimal solution of (1), by making use of generalized Fejér monotonicity techniques and of the Opial Lemma. The performed analysis allows us also to show the convergence of the objective function values to the optimal objective value of (1).

In the remaining of this section we recall some elements of convex analysis. For a function  $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$  we denote by  $\text{dom} f = \{x \in \mathcal{H} : f(x) < +\infty\}$  its effective domain and say that  $f$  is proper, if  $\text{dom} f \neq \emptyset$  and  $f(x) \neq -\infty$  for all  $x \in \mathcal{H}$ . Let  $f^* : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ ,  $f^*(u) = \sup_{x \in \mathcal{H}} \{\langle u, x \rangle - f(x)\}$  for all  $u \in \mathcal{H}$ , be the conjugate function of  $f$ . The subdifferential of  $f$  at  $x \in \mathcal{H}$ , with  $f(x) \in \mathbb{R}$ , is the set  $\partial f(x) := \{v \in \mathcal{H} : f(y) \geq f(x) + \langle v, y - x \rangle \forall y \in \mathcal{H}\}$ . We take by convention  $\partial f(x) := \emptyset$ , if  $f(x) \in \{\pm\infty\}$ . We also denote by  $\min f := \inf_{x \in \mathcal{H}} f(x)$  the optimal objective value of the function  $f$  and by  $\arg \min f := \{x \in \mathcal{H} : f(x) = \min f\}$  its set of global minima.

A convex and differentiable function  $g : \mathcal{H} \rightarrow \mathbb{R}$  has a Lipschitz continuous gradient with Lipschitz constant  $L_g > 0$ , if  $\|\nabla g(x) - \nabla g(y)\| \leq L_g \|x - y\|$  for all  $x, y \in \mathcal{H}$ . It is well-known (see, for instance, [12, Theorem 18.15]) that this is equivalent to  $\nabla g$  is  $\frac{1}{L_g}$ -cocoercive, namely,  $\langle x - y, \nabla g(x) - \nabla g(y) \rangle \geq \frac{1}{L_g} \|\nabla g(x) - \nabla g(y)\|^2$  for all  $x, y \in \mathcal{H}$ .

Let  $M \subseteq \mathcal{H}$  be a nonempty set. The indicator function of  $M$ ,  $\delta_M : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ , is the function which takes the value 0 on  $M$  and  $+\infty$  otherwise. The subdifferential of the indicator function is the normal cone of  $M$ , that is  $N_M(x) = \{u \in \mathcal{H} : \langle u, y - x \rangle \leq 0 \forall y \in M\}$ , if  $x \in M$  and  $N_M(x) = \emptyset$  for  $x \notin M$ . For  $x \in M$ , we have that  $u \in N_M(x)$  if and only if  $\sigma_M(u) = \langle u, x \rangle$ , where  $\sigma_M$  is the support function of  $M$ , defined by  $\sigma_M : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ ,  $\sigma_M(u) = \sup_{y \in M} \langle y, u \rangle$ . Finally,

$\text{ran}(N_M)$  denotes the range of the normal cone  $N_M$ , that is,  $p \in \text{ran}(N_M)$  if and only if there exists  $x \in M$  such that  $p \in N_M(x)$ .

## 2 Technical lemmas

The setting in which we will carry out the convergence analysis for Algorithm 1 is settled by the following hypotheses.

- Assumption 2** (I) *The subdifferential sum formula  $\partial(f + \delta_{\arg \min g}) = \partial f + N_{\arg \min g}$  holds;*  
 (II) *The objective function  $f + h$  is bounded from below;*  
 (III) *There exist positive constants  $\eta_0, a, b, K$  and  $c > 1$  such that for every  $n \geq 1$ :*

$$0 < a \leq \lambda_n \beta_n \leq b < \frac{2}{L_g(1 + \eta_0)^2}, \frac{L_h + \beta_n L_g}{2} + \frac{\alpha - 1}{\lambda_n} \leq -(1 + 2\alpha)K - c$$

and

$$\beta_{n+1} - \beta_n \leq K \frac{\eta_0}{1 + \eta_0} \lambda_{n+1} \beta_{n+1};$$

- (IV)  $\{\lambda_n\}_{n=1}^\infty \in \ell^2 \setminus \ell^1$  and  $\left(\frac{1}{\lambda_{n+1}} - \frac{1}{\lambda_n}\right) \alpha \leq 2$  for every  $n \geq 1$ ;  
 (V)  $\sum_{n=1}^\infty \lambda_n \beta_n \left[ g^* \left( \frac{p}{\beta_n} \right) - \sigma_{\arg \min g} \left( \frac{p}{\beta_n} \right) \right] < +\infty$  for every  $p \in \text{ran}(N_{\arg \min g})$ .

*Remark 1* For conditions which guarantee exact convex subdifferential sum formulas we refer to [12, 15, 16, 41]. One of these conditions, which is frequently fulfilled in applications, asks for the continuity of the function  $f$  and thus does not require any knowledge of the set of minimizers of  $g$ .

The assumption in (V) originates from the work of Attouch and Czarnecki [5]; we refer to [4–8, 11, 17–19, 24–27, 30, 37, 39] for other variants, generalizations to monotone operators and concrete examples for which this condition is satisfied (see also Remark 2).

The aim of the following three results is to derive a generalized Fejér-type inequality in the spirit of the one in the hypotheses of Lemma 4. This will be achieved in terms of the sequence  $(I_n)_{n \geq 1}$ , defined before Lemma 3, and which can be seen as a Lyapunov sequence equal to the sum of the objective function and a penalization of the function  $g$  both at the current iterate, and the distance of the current iterate to a fixed optimal solution.

**Lemma 1** *Let  $u \in \mathcal{S}$ . According to the first order optimality conditions there exist  $v \in \partial f(u)$  and  $p \in N_{\arg \min g}(u)$  such that  $0 = v + \nabla h(u) + p$ . Set*

$\varphi_n := \|x_n - u\|^2$  for every  $n \geq 1$ . Then for every  $n \geq 1$  and  $\eta > 0$  it holds

$$\begin{aligned} \varphi_{n+1} - \varphi_n - \alpha(\varphi_n - \varphi_{n-1}) + \lambda_n \beta_n \left[ \frac{2}{L_g(1+\eta)} - (1+\eta)\lambda_n \beta_n \right] \|\nabla g(x_n)\|^2 \\ + \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n) \leq 2\alpha \|x_n - x_{n-1}\|^2 + \|x_{n+1} - x_n\|^2 \\ + \frac{4(1+\eta)}{\eta} \lambda_n^2 \|\nabla h(u) + v\|^2 \\ + \lambda_n \left[ \frac{4(1+\eta)}{\eta} \lambda_n - \frac{2}{L_h} \right] \|\nabla h(x_n) - \nabla h(u)\|^2 \\ + \frac{\eta}{1+\eta} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right) \right] \end{aligned} \quad (3)$$

*Proof* Set  $y_n := x_n + \alpha(x_n - x_{n-1})$  for every  $n \geq 1$ . Since  $y_n - x_{n+1} - \lambda_n \nabla h(x_n) - \lambda_n \beta_n \nabla g(x_n) \in \lambda_n \partial f(x_{n+1})$  and  $v \in \partial f(u)$ , the monotonicity of  $\partial f$  guarantees that

$$\langle y_n - x_{n+1} - \lambda_n \nabla h(x_n) - \lambda_n \beta_n \nabla g(x_n) - \lambda_n v, x_{n+1} - u \rangle \geq 0 \quad \forall n \geq 1$$

or, equivalently,

$$2\langle y_n - x_{n+1}, u - x_{n+1} \rangle \leq 2\lambda_n \langle u - x_{n+1}, \nabla h(x_n) + \beta_n \nabla g(x_n) + v \rangle \quad \forall n \geq 1 \quad (4)$$

We notice that for every  $n \geq 0$

$$2\langle x_n - x_{n+1}, u - x_{n+1} \rangle = \varphi_{n+1} - \varphi_n + \|x_{n+1} - x_n\|^2 \quad (5)$$

and so for every  $n \geq 1$

$$\begin{aligned} 2\alpha \langle x_n - x_{n-1}, u - x_{n+1} \rangle &= 2\alpha \langle x_n - x_{n-1}, u - x_n \rangle + 2\alpha \langle x_n - x_{n-1}, x_n - x_{n+1} \rangle \\ &\geq \alpha (\|x_{n-1} - u\|^2 - \|x_n - u\|^2 - \|x_n - x_{n-1}\|^2) \\ &\quad + \alpha (-\|x_n - x_{n-1}\|^2 - \|x_{n+1} - x_n\|^2) \\ &= \alpha (\varphi_{n-1} - \varphi_n) - 2\alpha \|x_n - x_{n-1}\|^2 \\ &\quad - \alpha \|x_{n+1} - x_n\|^2. \end{aligned} \quad (6)$$

By employing (5) and (6) in inequality (4), we obtain for every  $n \geq 1$

$$\begin{aligned} \varphi_{n+1} - \varphi_n - \alpha(\varphi_n - \varphi_{n-1}) + (1-\alpha)\|x_{n+1} - x_n\|^2 - 2\alpha\|x_n - x_{n-1}\|^2 \\ \leq 2\lambda_n \langle u - x_{n+1}, \nabla h(x_n) + \beta_n \nabla g(x_n) + v \rangle \\ = 2\lambda_n \langle u - x_{n+1}, \beta_n \nabla g(x_n) \rangle + 2\lambda_n \langle u - x_{n+1}, \nabla h(x_n) + v \rangle \\ = 2\lambda_n \langle u - x_n, \beta_n \nabla g(x_n) \rangle + 2\lambda_n \langle x_n - x_{n+1}, \beta_n \nabla g(x_n) \rangle \\ + 2\lambda_n \langle u - x_{n+1}, \nabla h(x_n) + v \rangle. \end{aligned} \quad (7)$$

Next we evaluate the first two terms on the right-hand side in the above statement. Since  $\nabla g$  is  $\frac{1}{L_g}$ -cocoercive, we have

$$\langle \nabla g(x_n) - \nabla g(u), x_n - u \rangle \geq \frac{1}{L_g} \|\nabla g(x_n) - \nabla g(u)\|^2 \quad \forall n \geq 1,$$

and from here, since  $\nabla g(u) = 0$ ,

$$\langle \nabla g(x_n), u - x_n \rangle \leq -\frac{1}{L_g} \|\nabla g(x_n)\|^2 \quad \forall n \geq 1. \quad (8)$$

On the other hand, since  $g$  is convex and differentiable we have

$$0 = g(u) \geq g(x_n) + \langle \nabla g(x_n), u - x_n \rangle \quad \forall n \geq 1$$

or, equivalently,

$$\langle \nabla g(x_n), u - x_n \rangle \leq -g(x_n) \quad \forall n \geq 1. \quad (9)$$

From (8) and (9) we obtain for all  $n \geq 1$

$$2\lambda_n \beta_n \langle \nabla g(x_n), u - x_n \rangle \leq -\frac{2}{L_g(1+\eta)} \lambda_n \beta_n \|\nabla g(x_n)\|^2 - \frac{2\eta}{1+\eta} \lambda_n \beta_n g(x_n) \quad (10)$$

For the term  $2\lambda_n \beta_n \langle x_n - x_{n+1}, \nabla g(x_n) \rangle$  in (7) we have for all  $n \geq 1$  the following estimate

$$2\lambda_n \beta_n \langle x_n - x_{n+1}, \nabla g(x_n) \rangle \leq \frac{1}{1+\eta} \|x_{n+1} - x_n\|^2 + (1+\eta) \lambda_n^2 \beta_n^2 \|\nabla g(x_n)\|^2 \quad (11)$$

Employing the inequalities (10) and (11) in (7), we obtain for every  $n \geq 1$  that

$$\begin{aligned} & \varphi_{n+1} - \varphi_n - \alpha(\varphi_n - \varphi_{n-1}) + (1-\alpha)\|x_{n+1} - x_n\|^2 - 2\alpha\|x_n - x_{n-1}\|^2 \\ & \leq -\frac{2}{L_g(1+\eta)} \lambda_n \beta_n \|\nabla g(x_n)\|^2 - \frac{2\eta}{1+\eta} \lambda_n \beta_n g(x_n) \\ & \quad + \frac{1}{1+\eta} \|x_{n+1} - x_n\|^2 + (1+\eta) \lambda_n^2 \beta_n^2 \|\nabla g(x_n)\|^2 \\ & \quad + 2\lambda_n \langle u - x_{n+1}, \nabla h(x_n) + v \rangle, \end{aligned}$$

and further

$$\begin{aligned} & \varphi_{n+1} - \varphi_n - \alpha(\varphi_n - \varphi_{n-1}) + \left[ \frac{2}{L_g(1+\eta)} - (1+\eta) \lambda_n \beta_n \right] \lambda_n \beta_n \|\nabla g(x_n)\|^2 \\ & \leq 2\alpha\|x_n - x_{n-1}\|^2 - \frac{2\eta}{1+\eta} \lambda_n \beta_n g(x_n) \\ & \quad + \left[ \frac{1}{1+\eta} + \alpha - 1 \right] \|x_{n+1} - x_n\|^2 \\ & \quad + 2\lambda_n \langle u - x_{n+1}, \nabla h(x_n) + v \rangle. \end{aligned} \quad (12)$$

Not least,

$$\begin{aligned}
2\lambda_n \langle u - x_{n+1}, \nabla h(x_n) + v \rangle &= 2\lambda_n \langle u - x_n, \nabla h(x_n) + v \rangle \\
&\quad + 2\lambda_n \langle x_n - x_{n+1}, \nabla h(x_n) + v \rangle \\
&\leq 2\lambda_n \langle u - x_n, \nabla h(x_n) + v \rangle \\
&\quad + \frac{\eta}{2(1+\eta)} \|x_{n+1} - x_n\|^2 \\
&\quad + \frac{2(1+\eta)}{\eta} \lambda_n^2 \|\nabla h(x_n) + v\|^2 \\
&\leq 2\lambda_n \langle u - x_n, \nabla h(x_n) + v \rangle \\
&\quad + \frac{\eta}{2(1+\eta)} \|x_{n+1} - x_n\|^2 \\
&\quad + \frac{4(1+\eta)}{\eta} \lambda_n^2 \|\nabla h(x_n) - \nabla h(u)\|^2 \\
&\quad + \frac{4(1+\eta)}{\eta} \lambda_n^2 \|\nabla h(u) + v\|^2 \tag{13}
\end{aligned}$$

and by employing this estimate in (12), we deduce that for every  $n \geq 1$

$$\begin{aligned}
\varphi_{n+1} - \varphi_n - \alpha(\varphi_n - \varphi_{n-1}) + \lambda_n \beta_n \left[ \frac{2}{L_g(1+\eta)} - (1+\eta)\lambda_n \beta_n \right] \|\nabla g(x_n)\|^2 \\
+ \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n) \leq 2\alpha \|x_n - x_{n-1}\|^2 \\
+ \left[ \frac{1}{1+\eta} + \frac{\eta}{2(1+\eta)} + \alpha - 1 \right] \|x_{n+1} - x_n\|^2 \\
+ \frac{4(1+\eta)}{\eta} \lambda_n^2 \|\nabla h(x_n) - \nabla h(u)\|^2 \\
+ \frac{4(1+\eta)}{\eta} \lambda_n^2 \|\nabla h(u) + v\|^2 \\
- \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n) + 2\lambda_n \langle u - x_n, \nabla h(x_n) + v \rangle. \tag{14}
\end{aligned}$$

By using the  $\frac{1}{L_h}$ -cocoercivity of  $\nabla h$  we obtain for every  $n \geq 1$  that

$$\begin{aligned}
&2\lambda_n \langle u - x_n, \nabla h(x_n) + v \rangle - \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n) \\
&= 2\lambda_n \langle u - x_n, \nabla h(x_n) - \nabla h(u) \rangle + 2\lambda_n \langle u - x_n, \nabla h(u) + v \rangle \\
&\quad - \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n) \\
&\leq \frac{-2\lambda_n}{L_h} \|\nabla h(x_n) - \nabla h(u)\|^2 + 2\lambda_n \langle u - x_n, -p \rangle - \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n),
\end{aligned}$$

while, since  $p \in N_{\arg \min g}(u)$ , it holds

$$\begin{aligned}
& -2\lambda_n \langle u - x_n, p \rangle - \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n) \\
&= 2\lambda_n \langle x_n, p \rangle - \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n) - 2\lambda_n \langle u, p \rangle \\
&= \frac{\eta}{1+\eta} \lambda_n \beta_n \left[ \left\langle x_n, \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right\rangle - g(x_n) - \left\langle u, \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right\rangle \right] \\
&\leq \frac{\eta}{1+\eta} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right) - \left\langle u, \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right\rangle \right] \\
&= \frac{\eta}{1+\eta} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right) \right] \quad \forall n \geq 1.
\end{aligned}$$

By combining these two inequalities with (14), it follows for every  $n \geq 1$  that

$$\begin{aligned}
& \varphi_{n+1} - \varphi_n - \alpha(\varphi_n - \varphi_{n-1}) + \lambda_n \beta_n \left[ \frac{2}{L_g(1+\eta)} - (1+\eta)\lambda_n \beta_n \right] \|\nabla g(x_n)\|^2 \\
&+ \frac{\eta}{1+\eta} \lambda_n \beta_n g(x_n) \leq 2\alpha \|x_n - x_{n-1}\|^2 \\
&+ \left[ \frac{1}{1+\eta} + \frac{\eta}{2(1+\eta)} + \alpha - 1 \right] \|x_{n+1} - x_n\|^2 \\
&+ \frac{4(1+\eta)}{\eta} \lambda_n^2 \|\nabla h(u) + v\|^2 \\
&+ \lambda_n \left[ \frac{4(1+\eta)}{\eta} \lambda_n - \frac{2}{L_h} \right] \|\nabla h(x_n) - \nabla h(u)\|^2 \\
&+ \frac{\eta}{1+\eta} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta}{1+\eta} \beta_n} \right) \right].
\end{aligned}$$

Since  $\alpha \in [0, 1)$  and  $\eta > 0$ , it holds  $\frac{1}{1+\eta} + \frac{\eta}{2(1+\eta)} + \alpha - 1 < 1$ , which together with the inequality above lead to the conclusion.  $\square$

For simplicity, we will make use of the following notation:

$$\Omega_n(x_n) := f(x_n) + h(x_n) + \beta_n g(x_n) \quad \forall n \geq 1.$$

**Lemma 2** *For every  $n \geq 1$  it holds*

$$\begin{aligned}
\Omega_{n+1}(x_{n+1}) - \Omega_n(x_n) &\leq \left[ \frac{\alpha}{2\lambda_n} - \frac{1}{\lambda_n} + \frac{L_h + \beta_n L_g}{2} \right] \|x_{n+1} - x_n\|^2 \\
&+ \frac{\alpha}{2\lambda_n} \|x_n - x_{n-1}\|^2 + (\beta_{n+1} - \beta_n)g(x_{n+1}). \quad (15)
\end{aligned}$$



*Proof* Recall that for every  $n \geq 1$  we have  $\frac{y_n - x_{n+1}}{\lambda_n} - \nabla h(x_n) - \beta_n g(x_n) \in \partial f(x_{n+1})$ , which implies

$$f(x_n) \geq f(x_{n+1}) + \left\langle \frac{y_n - x_{n+1}}{\lambda_n} - \nabla h(x_n) - \beta_n g(x_n), x_n - x_{n+1} \right\rangle.$$

From here it follows that for every  $n \geq 1$  it holds

$$\begin{aligned} f(x_{n+1}) - f(x_n) &\leq -\frac{1}{\lambda_n} \|x_{n+1} - x_n\|^2 + \frac{\alpha}{\lambda_n} \langle x_n - x_{n-1}, x_{n+1} - x_n \rangle \\ &\quad + \langle \nabla h(x_n), x_n - x_{n+1} \rangle + \beta_n \langle \nabla g(x_n), x_n - x_{n+1} \rangle \\ &\leq \left[ \frac{\alpha}{2\lambda_n} - \frac{1}{\lambda_n} \right] \|x_{n+1} - x_n\|^2 + \frac{\alpha}{2\lambda_n} \|x_n - x_{n-1}\|^2 \\ &\quad + \langle \nabla h(x_n), x_n - x_{n+1} \rangle + \beta_n \langle \nabla g(x_n), x_n - x_{n+1} \rangle. \end{aligned} \quad (16)$$

From the Descent Lemma (see for example [12, Theorem 18.15]) we obtain for every  $n \geq 1$  that

$$g(x_{n+1}) \leq g(x_n) + \langle \nabla g(x_n), x_{n+1} - x_n \rangle + \frac{L_g}{2} \|x_{n+1} - x_n\|^2,$$

and

$$h(x_{n+1}) \leq h(x_n) + \langle \nabla h(x_n), x_{n+1} - x_n \rangle + \frac{L_h}{2} \|x_{n+1} - x_n\|^2.$$

By combining these relations with inequality (16) we finally obtain the inequality in the statement of the lemma.  $\square$

For the forthcoming statements we fix an element  $u \in \mathcal{S}$ . For a simpler formulation of these results we will use the following notation:

$$\begin{aligned} \Gamma_n &:= \Omega_n(x_n) - K \frac{\eta_0}{1 + \eta_0} \lambda_n \beta_n g(x_n) + K \varphi_n \\ &= f(x_n) + h(x_n) + \left( 1 - K \frac{\eta_0}{1 + \eta_0} \lambda_n \right) \beta_n g(x_n) + K \varphi_n \quad \forall n \geq 1. \end{aligned}$$

**Lemma 3** *Let  $u \in \mathcal{S}$ . According to the first order optimality conditions there exist  $v \in \partial f(u)$  and  $p \in N_{\arg \min g}(u)$  such that  $0 = v + \nabla h(u) + p$ . Then for*

every  $n \geq 2$  it holds

$$\begin{aligned}
& \Gamma_{n+1} - \Gamma_n - \alpha(\Gamma_n - \Gamma_{n-1}) + K\lambda_n\beta_n \left[ \frac{2}{L_g(1+\eta_0)} - (1+\eta_0)\lambda_n\beta_n \right] \|\nabla g(x_n)\|^2 \\
& \leq \left[ \frac{\alpha}{2\lambda_n} - \frac{1}{\lambda_n} + \frac{L_h + \beta_n L_g}{2} + K \right] \|x_{n+1} - x_n\|^2 \\
& \quad + \left[ \frac{\alpha}{2\lambda_n} + 2\alpha K \right] \|x_n - x_{n-1}\|^2 + \frac{4K(1+\eta_0)}{\eta_0} \lambda_n^2 \|\nabla h(u) + v\|^2 \\
& \quad + K \left[ \frac{4(1+\eta_0)}{\eta_0} \lambda_n^2 - \frac{2\lambda_n}{L_h} \right] \|\nabla h(x_n) - \nabla h(u)\|^2 \\
& \quad + \frac{K\eta_0}{1+\eta_0} \lambda_n\beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0}\beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0}\beta_n} \right) \right] \\
& \quad + \alpha(\Omega_{n-1}(x_{n-1}) - \Omega_n(x_n)) + \frac{\alpha K\eta_0}{1+\eta_0} (\lambda_n\beta_n g(x_n) - \lambda_{n-1}\beta_{n-1}g(x_{n-1})).
\end{aligned} \tag{17}$$

*Proof* We write (3) for  $\eta := \eta_0$ , multiply it by  $K$  and after combining the resulting inequality with (15) we obtain for every  $n \geq 1$  that

$$\begin{aligned}
& \Omega_{n+1}(x_{n+1}) + K\varphi_{n+1} - \Omega_n(x_n) - K\varphi_n + \frac{K\eta_0}{1+\eta_0} \lambda_n\beta_n g(x_n) \\
& \quad - \alpha(K\varphi_n - K\varphi_{n-1}) + K\lambda_n\beta_n \left[ \frac{2}{L_g(1+\eta_0)} - (1+\eta_0)\lambda_n\beta_n \right] \|\nabla g(x_n)\|^2 \\
& \leq \left[ \frac{\alpha}{2\lambda_n} - \frac{1}{\lambda_n} + \frac{L_h + \beta_n L_g}{2} + K \right] \|x_{n+1} - x_n\|^2 + \left[ \frac{\alpha}{2\lambda_n} + 2\alpha K \right] \|x_n - x_{n-1}\|^2 \\
& \quad + \frac{4K(1+\eta_0)}{\eta_0} \lambda_n^2 \|\nabla h(u) + v\|^2 + K \left[ \frac{4(1+\eta_0)}{\eta_0} \lambda_n^2 - \frac{2\lambda_n}{L_h} \right] \|\nabla h(x_n) - \nabla h(u)\|^2 \\
& \quad + (\beta_{n+1} - \beta_n)g(x_{n+1}) + \frac{K\eta_0}{1+\eta_0} \lambda_n\beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0}\beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0}\beta_n} \right) \right].
\end{aligned}$$

In view of Assumption 2(III) we deduce that

$$\begin{aligned}
& \Omega_{n+1}(x_{n+1}) + K\varphi_{n+1} - \Omega_n(x_n) - K\varphi_n + \frac{K\eta_0}{1+\eta_0} \lambda_n\beta_n g(x_n) \\
& \quad - \alpha(K\varphi_n - K\varphi_{n-1}) + K\lambda_n\beta_n \left[ \frac{2}{L_g(1+\eta_0)} - (1+\eta_0)\lambda_n\beta_n \right] \|\nabla g(x_n)\|^2 \\
& \leq \left[ \frac{\alpha}{2\lambda_n} - \frac{1}{\lambda_n} + \frac{L_h + \beta_n L_g}{2} + K \right] \|x_{n+1} - x_n\|^2 + \left[ \frac{\alpha}{2\lambda_n} + 2\alpha K \right] \|x_n - x_{n-1}\|^2 \\
& \quad + \frac{4K(1+\eta_0)}{\eta_0} \lambda_n^2 \|\nabla h(u) + v\|^2 + K \left[ \frac{4(1+\eta_0)}{\eta_0} \lambda_n^2 - \frac{2\lambda_n}{L_h} \right] \|\nabla h(x_n) - \nabla h(u)\|^2 \\
& \quad + \frac{K\eta_0}{1+\eta_0} \lambda_{n+1}\beta_{n+1}g(x_{n+1})
\end{aligned}$$

$$+ \frac{K\eta_0}{1+\eta_0} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) \right] \quad \forall n \geq 1$$

and further

$$\begin{aligned} & \Gamma_{n+1} - \Gamma_n - \alpha(K\varphi_n - K\varphi_{n-1}) \\ & + K\lambda_n \beta_n \left[ \frac{2}{L_g(1+\eta_0)} - (1+\eta_0)\lambda_n \beta_n \right] \|\nabla g(x_n)\|^2 \\ \leq & \left[ \frac{\alpha}{2\lambda_n} - \frac{1}{\lambda_n} + \frac{L_h + \beta_n L_g}{2} + K \right] \|x_{n+1} - x_n\|^2 + \left[ \frac{\alpha}{2\lambda_n} + 2\alpha K \right] \|x_n - x_{n-1}\|^2 \\ & + \frac{4K(1+\eta_0)}{\eta_0} \lambda_n^2 \|\nabla h(u) + v\|^2 + K \left[ \frac{4(1+\eta_0)}{\eta_0} \lambda_n^2 - \frac{2\lambda_n}{L_h} \right] \|\nabla h(x_n) - \nabla h(u)\|^2 \\ & + \frac{K\eta_0}{1+\eta_0} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) \right] \quad \forall n \geq 2. \end{aligned}$$

In order to obtain (17) we only have to add  $\alpha(\Omega_{n-1}(x_{n-1}) - \Omega_n(x_n))$  and  $\frac{\alpha K \eta_0}{1+\eta_0} (\lambda_n \beta_n g(x_n) - \lambda_{n-1} \beta_{n-1} g(x_{n-1}))$  to both sides of the above inequality.  $\square$

The following results is a very useful tool in the convergence analysis of inertial algorithms (see [1, 2, 20]).

**Lemma 4** *Let  $\{a_n\}_{n=0}^\infty$ ,  $\{b_n\}_{n=1}^\infty$  and  $\{c_n\}_{n=1}^\infty$  be real sequences and  $\alpha \in [0, 1)$  be a given real number. Assume that  $\{a_n\}_{n=1}^\infty$  is bounded from below,  $\{b_n\}_{n=1}^\infty$  is nonnegative and  $\sum_{n=1}^\infty c_n < +\infty$  such that*

$$a_{n+1} - a_n - \alpha(a_n - a_{n-1}) + b_n \leq c_n \quad \forall n \geq 1.$$

*Then the following statements hold:*

- (i)  $\sum_{n=1}^\infty [a_n - a_{n-1}]_+ < +\infty$ , where  $[t]_+ := \max\{t, 0\}$ ;
- (ii)  $\{a_n\}_{n=1}^\infty$  converges and  $\sum_{n=1}^\infty b_n < +\infty$ .

The results presented in Lemma 5 related to the convergence of the generated iterates and in Lemma 6 related to the convergence of the objective function values will be used in the next section in the proof of the main theorem in combination with the Opial Lemma stated in Lemma 7.

**Lemma 5** *Let  $u \in \mathcal{S}$ . According to the first order optimality conditions there exist  $v \in \partial f(u)$  and  $p \in N_{\arg \min g}(u)$  such that  $0 = v + \nabla h(u) + p$ . Then the following statements are true:*

- (i) *The sequence  $\{\Gamma_n\}_{n=1}^{+\infty}$  is bounded from below;*
- (ii)  $\sum_{n=1}^\infty \|x_{n+1} - x_n\|^2 < +\infty$ ;
- (iii)  $\lim_{n \rightarrow +\infty} \Gamma_n$  exists and  $\sum_{n=1}^{+\infty} \lambda_n \beta_n \|\nabla g(x_n)\|^2 < +\infty$ ;
- (iv)  $\lim_{n \rightarrow +\infty} \|x_n - u\|$  exists,  $\sum_{n=1}^\infty [\varphi_n - \varphi_{n-1}]_+ < +\infty$  and  $\sum_{n=1}^\infty \lambda_n \beta_n g(x_n) < +\infty$ ;
- (v)  $\lim_{n \rightarrow +\infty} \Omega_n(x_n)$  exists;

(vi)  $\lim_{n \rightarrow +\infty} g(x_n) = 0$  and every sequential weak cluster point of  $\{x_n\}_{n=1}^{\infty}$  lies in  $\arg \min g$ .

*Proof* (i) According to Assumption 2(III) we have that

$$\frac{L_h + \beta_n L_g}{2} + \frac{\alpha - 1}{\lambda_n} \leq -(1 + 2\alpha)K - c,$$

which implies that  $1 \geq K\lambda_n$  and further

$$\left(1 - \frac{K\eta_0}{1 + \eta_0}\lambda_n\right) \beta_n g(x_n) \geq 0 \quad \forall n \geq 1. \quad (18)$$

By using the definition of  $\Gamma_n$  and Assumption 2(II), we easily derive that  $\{\Gamma_n\}_{n=1}^{\infty}$  is bounded from below.

(ii) For every  $n \geq 2$  we set

$$\mu_n := \Gamma_n - \alpha\Gamma_{n-1} + \left(\frac{1}{2\lambda_n} + 2K\right) \alpha \|x_n - x_{n-1}\|^2$$

and

$$\omega_n := \alpha(\Omega_{n-1}(x_{n-1}) - \Omega_n(x_n)) + \frac{\alpha K \eta_0}{1 + \eta_0} (\lambda_n \beta_n g(x_n) - \lambda_{n-1} \beta_{n-1} g(x_{n-1})).$$

For a fixed natural number  $N_0 \geq 2$  it holds

$$\begin{aligned} \frac{1}{\alpha} \sum_{n=2}^{N_0} \omega_n &= f(x_1) + h(x_1) + \left(1 - \frac{K\eta_0}{1 + \eta_0}\lambda_1\right) \beta_1 g(x_1) - f(x_{N_0}) - h(x_{N_0}) \\ &\quad - \left(1 - \frac{K\eta_0}{1 + \eta_0}\lambda_{N_0}\right) \beta_{N_0} g(x_{N_0}). \end{aligned}$$

Since  $f + h$  is bounded from below and relation (18) is true, we obtain that  $\{\omega_n\}_{n=1}^{\infty}$  is summable.

For every  $n \geq 1$  we set

$$\delta_n := \left(\frac{1}{2\lambda_n} + 2K\right) \alpha + c.$$

Consequently, according Assumption 2(III) it follows

$$\frac{L_h + \beta_n L_g}{2} + \frac{\alpha}{2\lambda_n} - \frac{1}{\lambda_n} + K \leq -\delta_n \quad \forall n \geq 1. \quad (19)$$

Further, for every  $n \geq 1$  it holds

$$-\delta_n + \alpha \left(\frac{1}{2\lambda_{n+1}} + 2K\right) = \frac{\alpha}{2} \left(\frac{1}{\lambda_{n+1}} - \frac{1}{\lambda_n}\right) - c,$$

which, together with Assumption 2(IV), implies that

$$-\delta_n + \alpha \left(\frac{1}{2\lambda_{n+1}} + 2K\right) \leq 1 - c \quad (20)$$

and so

$$\delta_{n+1} \leq \delta_n + 1. \quad (21)$$

On the other hand, by Assumption 2(III), we also have for every  $n \geq 1$

$$0 < \frac{2}{L_g(1+\eta_0)} - (1+\eta_0)b \leq \frac{2}{L_g(1+\eta_0)} - (1+\eta_0)\lambda_n\beta_n \quad (22)$$

and so

$$K\lambda_n\beta_n \left[ \frac{2}{L_g(1+\eta_0)} - (1+\eta_0)\lambda_n\beta_n \right] \geq 0.$$

By employing the last inequality in Lemma 3 we obtain for every  $n \geq 2$

$$\begin{aligned} \mu_{n+1} - \mu_n &= \Gamma_{n+1} - \Gamma_n - \alpha(\Gamma_n - \Gamma_{n-1}) + \alpha \left( \frac{1}{2\lambda_{n+1}} + 2K \right) \|x_{n+1} - x_n\|^2 \\ &\quad - \alpha \left( \frac{1}{2\lambda_n} + 2K \right) \|x_n - x_{n-1}\|^2 \\ &\leq \left[ \frac{\alpha}{2\lambda_n} - \frac{1}{\lambda_n} + \frac{L_h + \beta_n L_g}{2} + K + \alpha \left( \frac{1}{2\lambda_{n+1}} + 2K \right) \right] \|x_{n+1} - x_n\|^2 \\ &\quad + \frac{4K(1+\eta_0)}{\eta_0} \lambda_n^2 \|\nabla h(u) + v\|^2 \\ &\quad + K \left[ \frac{4(1+\eta_0)}{\eta_0} \lambda_n^2 - \frac{2\lambda_n}{L_h} \right] \|\nabla h(x_n) - \nabla h(u)\|^2 \\ &\quad + \frac{K\eta_0}{1+\eta_0} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) \right] + \omega_n \\ &\leq (1-c) \|x_{n+1} - x_n\|^2 + \frac{4K(1+\eta_0)}{\eta_0} \lambda_n^2 \|\nabla h(u) + v\|^2 \\ &\quad + K \left[ \frac{4(1+\eta_0)}{\eta_0} \lambda_n^2 - \frac{2\lambda_n}{L_h} \right] \|\nabla h(x_n) - \nabla h(u)\|^2 \\ &\quad + \frac{K\eta_0}{1+\eta_0} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) \right] + \omega_n, \end{aligned}$$

where for the last inequality we use (19) and (20). Since  $\lambda_n \rightarrow 0$  as  $n \rightarrow +\infty$ , there exists  $N_1 \in \mathbb{N}$  such that for every  $n \geq N_1$  we have  $\frac{4(1+\eta_0)}{\eta_0} \lambda_n - \frac{2}{L_h} < 0$ . This implies that for every  $n \geq N_1$

$$\begin{aligned} \mu_{n+1} - \mu_n &\leq (1-c) \|x_{n+1} - x_n\|^2 + \frac{4K(1+\eta_0)}{\eta_0} \lambda_n^2 \|\nabla h(u) + v\|^2 \\ &\quad + \frac{K\eta_0}{1+\eta_0} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) \right] + \omega_n. \end{aligned}$$

Summing up this inequality for  $n = N_1, \dots, N_2$ , where  $N_2$  is a natural number with  $N_2 \geq N_1$ , we obtain that

$$\begin{aligned} \mu_{N_2+1} - \mu_{N_1} &\leq (1-c) \sum_{n=N_1}^{N_2} \|x_{n+1} - x_n\|^2 + \frac{4K(1+\eta_0)}{\eta_0} \|\nabla h(u) + v\|^2 \sum_{n=N_1}^{N_2} \lambda_n^2 \\ &\quad + \frac{K\eta_0}{1+\eta_0} \sum_{n=N_1}^{N_2} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) \right] \\ &\quad + \sum_{n=N_1}^{N_2} \omega_n. \end{aligned} \tag{23}$$

This means that  $\{\mu_n\}_{n=2}^\infty$  is bounded from above (we take into account that  $c > 1$ ). Let  $M$  be a positive upper bound of  $\{\mu_n\}_{n=2}^\infty$ . Observing that  $\Gamma_{n+1} - \alpha\Gamma_n \leq \mu_{n+1} \leq M$ , thus  $\Gamma_{n+1} \leq \alpha\Gamma_n + M$ , for every  $n \geq N_1$ , we obtain

$$\Gamma_n \leq \alpha^{n-N_1} \Gamma_{N_1} + M \sum_{k=1}^{n-N_1} \alpha^{k-1} \leq \alpha^{n-N_1} \Gamma_{N_1} + \frac{M}{1-\alpha} \quad \forall n \geq N_1 + 1.$$

Since  $\{\Gamma_n\}_{n=1}^\infty$  is bounded from below, there exists  $C \in \mathbb{R}$  such that

$$\begin{aligned} -\mu_{N_2+1} &= -\Gamma_{N_2+1} + \alpha\Gamma_{N_2} - \left( \frac{1}{2\lambda_{N_2+1}} + 2K \right) \alpha \|x_{N_2+1} - x_{N_2}\|^2 \\ &\leq \alpha\Gamma_{N_2} + C \leq \alpha^{N_2-N_1+1} \Gamma_{N_1} + \frac{M\alpha}{1-\alpha} + C \quad \forall N_2 \geq N_1. \end{aligned}$$

Thus, from the inequality (23), by taking into account that  $c > 1$ , we deduce that

$$\sum_{n=1}^{+\infty} \|x_{n+1} - x_n\|^2 < +\infty.$$

(iii) From Lemma 3, by using (19), (21) and (22), we obtain

$$\begin{aligned} \Gamma_{n+1} - \Gamma_n - \alpha(\Gamma_n - \Gamma_{n-1}) &+ K\lambda_n\beta_n \left[ \frac{2}{L_g(1+\eta_0)} - (1+\eta_0)b \right] \|\nabla g(x_n)\|^2 \\ &\leq -\delta_n \|x_{n+1} - x_n\|^2 + \delta_{n-1} \|x_n - x_{n-1}\|^2 + \frac{4K(1+\eta_0)}{\eta_0} \lambda_n^2 \|\nabla h(u) + v\|^2 \\ &\quad + \frac{K\eta_0}{1+\eta_0} \lambda_n \beta_n \left[ g^* \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) - \sigma_{\arg \min g} \left( \frac{2p}{\frac{\eta_0}{1+\eta_0} \beta_n} \right) \right] + \omega_n \quad \forall n \geq N_1. \end{aligned}$$

Since  $\{\Gamma_n\}_{n=1}^\infty$  is bounded from below, by using Lemma 4, it follows that (iii) is true.

(iv) The statement follows from Lemma 1 and Lemma 4.

(v) Thanks to (iii), (iv) and  $\Gamma_n = \Omega_n(x_n) - K\frac{\eta_0}{1+\eta_0} \lambda_n \beta_n g(x_n) + K\varphi_n$  for every  $n \geq 1$ , we obtain that  $\lim_{n \rightarrow +\infty} \Omega_n(x_n)$  exists.

(vi) Since  $\lambda_n \beta_n \geq a > 0$  for every  $n \geq 1$ , we have  $\sum_{n=1}^{+\infty} g(x_n) < +\infty$  and so  $\lim_{n \rightarrow +\infty} g(x_n) = 0$ .

Finally, let  $y \in \mathcal{H}$  be a sequential weak cluster point of  $\{x_n\}_{n=1}^{\infty}$  and  $\{x_{n_j}\}_{j=1}^{\infty}$  be a subsequence of  $\{x_n\}_{n=1}^{\infty}$  such that  $x_{n_j}$  weakly converges to  $y$  as  $j \rightarrow +\infty$ . Since  $g$  is weakly lower semicontinuous, we obtain that

$$g(y) \leq \liminf_{j \rightarrow +\infty} g(x_{n_j}) \leq \lim_{n \rightarrow +\infty} g(x_n) = 0,$$

which means that  $y \in \arg \min g$ .  $\square$

**Lemma 6** *Let  $u \in \mathcal{S}$ . Then we have*

$$-\infty \leq \sum_{n=1}^{\infty} \lambda_n [\Omega_{n+1}(x_{n+1}) - (f(u) + h(u))] < +\infty.$$

*Proof* For every  $n \geq 1$  we have that

$$\begin{aligned} \Omega_{n+1}(x_{n+1}) - (f(u) + h(u)) &= f(x_{n+1}) + h(x_n) + \beta_n g(x_n) - (f(u) + h(u)) \\ &\quad + (h(x_{n+1}) - h(x_n)) + (\beta_{n+1} - \beta_n)g(x_{n+1}) \\ &\quad + \beta_n(g(x_{n+1}) - g(x_n)) \\ &\leq f(x_{n+1}) + h(x_n) + \beta_n g(x_n) - (f(u) + h(u)) \\ &\quad + (h(x_{n+1}) - h(x_n)) + \frac{K\eta_0}{1 + \eta_0} \lambda_{n+1} \beta_{n+1} g(x_{n+1}) \\ &\quad + \beta_n(g(x_{n+1}) - g(x_n)) \end{aligned}$$

and so

$$\begin{aligned} \lambda_n [\Omega_{n+1}(x_{n+1}) - (f(u) + h(u))] &\leq \lambda_n [f(x_{n+1}) + h(x_n) + \beta_n g(x_n) \\ &\quad - (f(u) + h(u))] \\ &\quad + \lambda_n [(h(x_{n+1}) - h(x_n))] \\ &\quad + \frac{K\eta_0}{1 + \eta_0} \lambda_n \lambda_{n+1} \beta_{n+1} g(x_{n+1}) \\ &\quad + \lambda_n \beta_n [g(x_{n+1}) - g(x_n)]. \end{aligned}$$

According to the Descent Lemma we have for every  $n \geq 1$

$$\begin{aligned} \lambda_n [h(x_{n+1}) - h(x_n)] &\leq \lambda_n \langle \nabla h(x_n), x_{n+1} - x_n \rangle + \frac{\lambda_n L_h}{2} \|x_{n+1} - x_n\|^2 \\ &\leq \frac{\lambda_n^2}{2} \|\nabla h(x_n)\|^2 + \frac{1 + \lambda_n L_h}{2} \|x_{n+1} - x_n\|^2 \end{aligned}$$

and

$$\begin{aligned} \lambda_n \beta_n [g(x_{n+1}) - g(x_n)] &\leq \lambda_n \beta_n \langle \nabla g(x_n), x_{n+1} - x_n \rangle + \frac{\lambda_n \beta_n L_g}{2} \|x_{n+1} - x_n\|^2 \\ &\leq \frac{\lambda_n^2 \beta_n^2}{2} \|\nabla g(x_n)\|^2 + \frac{1 + \lambda_n \beta_n L_g}{2} \|x_{n+1} - x_n\|^2, \end{aligned}$$

which give rise to the following estimate

$$\begin{aligned}
\lambda_n[\Omega_{n+1}(x_{n+1}) - (f(u) + h(u))] &\leq \lambda_n[f(x_{n+1}) + h(x_n) + \beta_n g(x_n) \\
&\quad - (f(u) + h(u))] \\
&\quad + \frac{\lambda_n^2}{2} \|\nabla h(x_n)\|^2 \\
&\quad + \frac{2 + \lambda_n(L_h + \beta_n L_g)}{2} \|x_{n+1} - x_n\|^2 \\
&\quad + \frac{K\eta_0}{1 + \eta_0} \lambda_n \lambda_{n+1} \beta_{n+1} g(x_{n+1}) \\
&\quad + \frac{\lambda_n^2 \beta_n^2}{2} \|\nabla g(x_n)\|^2 \quad \forall n \geq 1. \tag{24}
\end{aligned}$$

Further, we notice that for every  $n \geq 1$

$$f(x_{n+1}) - f(u) \leq \left\langle \frac{y_n - x_{n+1}}{\lambda_n} - \nabla h(x_n) - \beta_n \nabla g(x_n), x_{n+1} - u \right\rangle$$

or, equivalently,

$$\begin{aligned}
\left\langle \frac{y_n - x_{n+1}}{\lambda_n}, u - x_{n+1} \right\rangle - \langle \nabla h(x_n), u - x_{n+1} \rangle - \beta_n \langle \nabla g(x_n), u - x_{n+1} \rangle \\
\leq f(u) - f(x_{n+1}). \tag{25}
\end{aligned}$$

On the other hand, since  $g(u) = 0$ , we have for every  $n \geq 1$

$$0 = g(u) \geq g(x_n) + \langle \nabla g(x_n), u - x_n \rangle$$

or, equivalently,

$$-\beta_n \langle \nabla g(x_n), u - x_{n+1} \rangle \geq \beta_n g(x_n) + \beta_n \langle \nabla g(x_n), x_{n+1} - x_n \rangle. \tag{26}$$

Similarly, we have for every  $n \geq 1$

$$h(u) \geq h(x_n) + \langle \nabla h(x_n), u - x_n \rangle,$$

which implies that

$$-\langle \nabla h(x_n), u - x_{n+1} \rangle \geq h(x_n) - h(u) + \langle \nabla h(x_n), x_{n+1} - x_n \rangle. \tag{27}$$

By summing up the inequalities in (25)-(27), we obtain for every  $n \geq 1$

$$\begin{aligned}
2\lambda_n[f(x_{n+1}) + h(x_n) + \beta_n g(x_n) - (f(u) + h(u))] &\leq 2\langle y_n - x_{n+1}, x_{n+1} - u \rangle \\
&\quad + 2\lambda_n \beta_n \langle \nabla g(x_n), x_n - x_{n+1} \rangle \\
&\quad + 2\lambda_n \langle \nabla h(x_n), x_n - x_{n+1} \rangle \\
&\leq 2\langle y_n - x_{n+1}, x_{n+1} - u \rangle \\
&\quad + \lambda_n^2 \beta_n^2 \|\nabla g(x_n)\|^2 + 2\|x_n - x_{n+1}\|^2 \\
&\quad + \lambda_n^2 \|\nabla h(x_n)\|^2. \tag{28}
\end{aligned}$$



Not least, according to (5) and (6) we obtain for every  $n \geq 1$

$$\begin{aligned} 2\langle y_n - x_{n+1}, x_{n+1} - u \rangle &= 2\langle x_n - x_{n+1}, x_{n+1} - u \rangle + 2\alpha\langle x_n - x_{n-1}, x_{n+1} - u \rangle \\ &\leq \varphi_n - \varphi_{n+1} + \alpha(\varphi_n - \varphi_{n-1}) \\ &\quad + 2\alpha\|x_n - x_{n-1}\|^2 + (\alpha - 1)\|x_{n+1} - x_n\|^2 \\ &\leq \varphi_n - \varphi_{n+1} + \alpha(\varphi_n - \varphi_{n-1}) + 2\alpha\|x_n - x_{n-1}\|^2, \end{aligned}$$

which, combined with (28) and Lemma 5(iv), implies that

$$\sum_{n=1}^{\infty} \lambda_n [f(x_{n+1}) + h(x_n) + \beta_n g(x_n) - (f(u) + h(u))] < +\infty.$$

The conclusion follows by taking into account (24).  $\square$

### 3 Convergence of the iterates and of the objective function values

In this section we will prove the main result of this paper. This addresses the convergence of both the sequence of iterates  $\{x_n\}_{n=1}^{\infty}$  generated by Algorithm 1 and of the sequence of objective values  $\{f(x_n) + h(x_n)\}_{n=1}^{\infty}$ .

The Opial Lemma, which we state next and for which we refer to [12, Lemma 2.39], will play a crucial role in the convergence analysis.

**Lemma 7** *Let  $\mathcal{H}$  be a real Hilbert space,  $C \subseteq \mathcal{H}$  a nonempty set and  $\{x_n\}_{n=1}^{\infty}$  a given sequence such that:*

- (i) *For every  $z \in C$ ,  $\lim_{n \rightarrow +\infty} \|x_n - z\|$  exists.*
- (ii) *Every sequential weak cluster point of  $\{x_n\}_{n=1}^{\infty}$  lies in  $C$ .*

*Then the sequence  $\{x_n\}_{n=1}^{\infty}$  converges weakly to a point in  $C$ .*

**Theorem 3** *Let  $\{x_n\}_{n=1}^{\infty}$  be the sequence generated by Algorithm 1. Then:*

- (i) *the sequence  $\{x_n\}_{n=1}^{\infty}$  converges weakly to a point in  $\mathcal{S}$ ;*
- (ii) *the sequence  $\{f(x_n) + h(x_n)\}_{n=1}^{\infty}$  converges to the optimal objective value of the optimization problem (1).*

*Proof* (i) We know that  $\lim_{n \rightarrow +\infty} \|x_n - u\|$  exists for all  $u \in \mathcal{S}$  (see Lemma 5(iv)), hence in view of the Opial Lemma it is sufficient to show that all sequential weak cluster points of  $\{x_n\}_{n=1}^{\infty}$  are in  $\mathcal{S}$ . Since  $\{\lambda_n\}_{n=1}^{\infty} \notin \ell^1$  and  $\lim_{n \rightarrow +\infty} \Omega_n(x_n)$  exists, from Lemma 6 we obtain that

$$\lim_{n \rightarrow +\infty} \Omega_n(x_n) \leq f(u) + h(u) \quad \forall u \in \mathcal{S}.$$

Let  $x^* \in \mathcal{H}$  be a sequential weak cluster point of  $\{x_n\}_{n=1}^{\infty}$  and  $\{x_{n_k}\}_{k=1}^{\infty}$  be a subsequence of  $\{x_n\}_{n=1}^{\infty}$  such that  $x_{n_k}$  converges weakly to  $x^*$  as  $k \rightarrow +\infty$ .

From here, by Lemma 5(vi), we obtain that  $x^* \in \arg \min g$ . Take an arbitrary  $u \in \mathcal{S}$ . The weak lower semicontinuity of  $f$  and  $h$  implies that

$$\begin{aligned} f(x^*) + h(x^*) &\leq \liminf_{k \rightarrow +\infty} f(x_{n_k}) + \liminf_{k \rightarrow +\infty} h(x_{n_k}) \\ &\leq \liminf_{k \rightarrow +\infty} [f(x_{n_k}) + h(x_{n_k})] \\ &\leq \lim_{n \rightarrow +\infty} \Omega_n(x_n) \leq f(u) + h(u) \\ &= \min\{f(x) + h(x) : x \in \arg \min g\} \end{aligned}$$

which means that  $x^* \in \mathcal{S}$ .

(ii) The statement is a direct consequence of the above inequalities.  $\square$

We close the paper by a remark which discusses the fulfillment of the conditions stated in Assumption 2.

*Remark 2* We chose

$$\eta_0 \in (0, +\infty), c \in (1, +\infty), q \in \left(\frac{1}{2}, 1\right), \alpha \in \left(0, 1 - \frac{1}{2(1 + \eta_0)^2}\right)$$

and

$$\gamma \in \left(\frac{1}{L_g(1 - \alpha)(1 + \eta_0)^2}, \min\left\{\frac{2}{L_g}, \frac{3}{L_g(1 - \alpha)(1 + \eta_0)^2}\right\}\right).$$

We set

$$K := \frac{2(1 + \eta_0)}{\alpha\eta_0},$$

$$\beta_n := \frac{\gamma[L_h + 2((1 + 2\alpha)K + c)]}{2 - \gamma L_g} + \left[\frac{(1 - \alpha)\gamma L_g(1 + \eta_0)^2 - 1}{L_g(1 + \eta_0)^2}\right] \frac{K\eta_0}{1 + \eta_0} n^q,$$

and

$$\lambda_n := \frac{(1 - \alpha)\gamma}{\beta_n} - \frac{1}{\beta_n L_g(1 + \eta_0)^2},$$

for every  $n \geq 1$ .

(i) By taking  $a := (1 - \alpha)\gamma - \frac{1}{L_g(1 + \eta_0)^2}$  and  $b > 0$  such that  $a < b < \frac{2}{L_g(1 + \eta_0)^2}$ , we have

$$0 < a = \lambda_n \beta_n < b < \frac{2}{L_g(1 + \eta_0)^2} \quad \forall n \geq 1.$$

(ii) Since  $\beta_n \geq \frac{\gamma[L_h + 2((1 + 2\alpha)K + c)]}{2 - \gamma L_g}$ , we have  $\frac{L_h + \beta_n L_g}{2} - \frac{\beta_n}{\gamma} \leq -(1 + 2\alpha)K - c$  for every  $n \geq 1$ . On the other hand, since  $\frac{\beta_n}{\gamma} \leq \frac{(1 - \alpha)}{\lambda_n}$ , we have that  $\frac{L_h + \beta_n L_g}{2} + \frac{\alpha - 1}{\lambda_n} \leq -(1 + 2\alpha)K - c$  for every  $n \geq 1$ .

(iii) For every  $n \geq 1$  we also have

$$\begin{aligned}\beta_{n+1} - \beta_n &= \left[ \frac{(1-\alpha)\gamma L_g(1+\eta_0)^2 - 1}{L_g(1+\eta_0)^2} \right] \frac{K\eta_0}{1+\eta_0} ((n+1)^q - n^q) \\ &\leq \left[ \frac{(1-\alpha)\gamma L_g(1+\eta_0)^2 - 1}{L_g(1+\eta_0)^2} \right] \frac{K\eta_0}{1+\eta_0} \\ &= \frac{K\eta_0}{1+\eta_0} \lambda_{n+1} \beta_{n+1}.\end{aligned}$$

(iv) From (iii) it follows that for every  $n \geq 1$

$$\frac{1}{\lambda_{n+1}} - \frac{1}{\lambda_n} = (\beta_{n+1} - \beta_n) \left[ \frac{L_g(1+\eta_0)^2}{(1-\alpha)\gamma L_g(1+\eta_0)^2 - 1} \right] \leq \frac{K\eta_0}{1+\eta_0} = \frac{2}{\alpha}.$$

(v) Due to the fact that  $q \in (\frac{1}{2}, 1)$ , we have  $\sum_{n=1}^{+\infty} \frac{1}{\beta_n} = +\infty$  and  $\sum_{n=1}^{+\infty} \frac{1}{\beta_n^2} < +\infty$ . Consequently,  $\{\lambda_n\}_{n=1}^{\infty} \in \ell^2 \setminus \ell^1$ .

(vi) Since  $g \leq \delta_{\arg \min g}$ , it holds  $g^* \geq (\delta_{\arg \min g})^* = \sigma_{\arg \min g}$  and so  $g^* - \sigma_{\arg \min g} \geq 0$ . For a function  $g : \mathcal{H} \rightarrow \mathbb{R}$  fulfilling  $g \geq \frac{a}{2} \text{dist}^2(\cdot, \arg \min g)$  where  $a > 0$ , it holds  $g^*(x) - \sigma_{\arg \min g}(x) \leq \frac{1}{2a} \|x\|^2$  for every  $x \in \mathcal{H}$ . Thus, for every  $n \geq 1$ ,

$$\lambda_n \beta_n \left[ g^* \left( \frac{p}{\beta_n} \right) - \sigma_{\arg \min g} \left( \frac{p}{\beta_n} \right) \right] \leq \frac{\lambda_n}{2a\beta_n} \|p\|^2 \quad \forall p \in \text{ran}(N_{\arg \min g}).$$

Since  $\sum_{n=1}^{\infty} \frac{1}{\beta_n^2} < +\infty$ , from here it follows that

$$\sum_{n=1}^{\infty} \lambda_n \beta_n \left[ g^* \left( \frac{p}{\beta_n} \right) - \sigma_{\arg \min g} \left( \frac{p}{\beta_n} \right) \right] < +\infty.$$

## 4 Conclusions

We investigate the weak (non-ergodic) convergence of an inertial proximal-gradient method with penalization terms in connection with the solving of a bilevel optimization problem, having as objective the sum of a convex nonsmooth with a convex smooth function, and as constrained set the set of minimizers of another convex and differentiable function. The techniques of the proof combine tools specific to inertial algorithms [3] and penalty type methods [5, 8]. We show the convergence of both generated iterates and objective function values.

**Acknowledgements** The authors are thankful to two anonymous referees and the Guest Editor Regina Burachik for comments and remarks which improved the quality of the presentation.

## References

1. F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM Journal on Control and Optimization 38(4), 1102–1119 (2000)
2. F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM Journal on Optimization 14(3), 773–782 (2004)
3. F. Alvarez, H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis 9, 3–11 (2001)
4. H. Attouch, A. Cabot, M.-O. Czarnecki, *Asymptotic behavior of nonautonomous monotone and subgradient evolution equations*, to appear in Transactions of the American Mathematical Society, DOI: <https://doi.org/10.1090/tran/6965>, arXiv:1601.00767 (2016)
5. H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, Journal of Differential Equations 248(6), 1315–1344 (2010)
6. H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects*, Journal of Differential Equations 262(3), 2745–2770 (2010)
7. 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)
8. 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)
9. H. Attouch, J. Peypouquet, *The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than  $1/k^2$* , SIAM Journal on Optimization 26(3), 1824–1834 (2016)
10. 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)
11. S. Banert, R.I. Boț, *Backward penalty schemes for monotone inclusion problems*, Journal of Optimization Theory and Applications 166(3), 930–948 (2015)
12. H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York (2011)
13. A. Beck, M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences 2(1), 183–202 (2009)
14. D.P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Cambridge, MA (1999)
15. J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge (2010)
16. R.I. Boț, *Conjugate Duality in Convex Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 637, Springer, Berlin Heidelberg (2010)
17. R.I. Boț, E.R. Csetnek, *Forward-backward and Tseng’s type penalty schemes for monotone inclusion problems*, Set-Valued and Variational Analysis 22, 313–331 (2014)
18. R.I. Boț, 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)
19. R.I. Boț, E.R. Csetnek, *Levenberg-Marquardt dynamics associated to variational inequalities*, to appear in Set-Valued and Variational Analysis, DOI 10.1007/s11228-017-0409-8, arXiv:1603.04460 (2016)
20. R.I. Boț, E.R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numerical Algorithms 71, 519–540 (2016)
21. R.I. Boț, E.R. Csetnek, *An inertial alternating direction method of multipliers*, Minimax Theory and its Applications 1(1), 29–49 (2016)
22. R.I. Boț, E.R. Csetnek, *A hybrid proximal-extragradient algorithm with inertial effects*, Numerical Functional Analysis and Optimization 36(8), 951–963 (2015)
23. R.I. Boț, E.R. Csetnek, *An inertial Tseng’s type proximal algorithm for nonsmooth and nonconvex optimization problems*, Journal of Optimization Theory and Applications 171(2), 600–616 (2016)

24. 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, 1688–1700 (2016)
25. R.I. Boţ, E.R. Csetnek, *Penalty schemes with inertial effects for monotone inclusion problems*, Optimization 66(6), 965–982 (2017)
26. R.I. Boţ, E.R. Csetnek, *Second order dynamical systems associated to variational inequalities*, Applicable Analysis 96(5), 799–809 (2017)
27. R.I. Boţ, E.R. Csetnek, *A second order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities*, arXiv:1608.04137 (2016)
28. R.I. Boţ, E.R. Csetnek, C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Applied Mathematics and Computation 256, 472–487 (2015)
29. 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)
30. R.I. Boţ, E.R. Csetnek, N. Nimana, *Gradient-type penalty method with inertial effects for solving constrained convex optimization problems with smooth data*, to appear in Optimization Letters, DOI 10.1007/s11590-017-1158-1
31. A. Cabot, P. Frankel, *Asymptotics for some proximal-like method involving inertia and memory aspects*, Set-Valued and Variational Analysis 19, 59–74 (2011)
32. C. Chen, R.H. Chan, S. MA, J. Yang, *Inertial proximal ADMM for linearly constrained separable convex optimization*, SIAM Journal on Imaging Sciences 8(4), 2239–2267 (2015)
33. C. Chen, S. MA, J. Yang, *A general inertial proximal point algorithm for mixed variational inequality problem*, SIAM Journal on Optimization 25(4), 2120–2142 (2015)
34. P.-E. Maingé, *Convergence theorems for inertial KM-type algorithms*, Journal of Computational and Applied Mathematics 219, 223–236 (2008)
35. P.-E. Maingé, A. Moudafi, *Convergence of new inertial proximal methods for dc programming*, SIAM Journal on Optimization 19(1), 397–413 (2008)
36. A. Moudafi, M. Oliny, *Convergence of a splitting inertial proximal method for monotone operators*, Journal of Computational and Applied Mathematics 155, 447–454 (2003)
37. N. Noun, J. Peyrouquet, *Forward-backward penalty scheme for constrained convex minimization without inf-compactness*, Journal of Optimization Theory and Applications, 158(3), 787–795 (2013)
38. P. Ochs, Y. Chen, T. Brox, T. Pock, *iPiano: Inertial proximal algorithm for non-convex optimization*, SIAM Journal on Imaging Sciences 7(2), 1388–1419 (2014)
39. J. Peyrouquet, *Coupling the gradient method with a general exterior penalization scheme for convex minimization*, Journal of Optimization Theory and Applications 153(1), 123–138 (2012)
40. B.T. Polyak, *Introduction to Optimization*, (Translated from the Russian) Translations Series in Mathematics and Engineering, Optimization Software, Inc., Publications Division, New York (1987)
41. C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, Singapore (2002)