

# A gradient type algorithm with backward inertial steps for a nonconvex minimization

Cristian Daniel Alecsa\*

Szilárd Csaba László†

Adrian Viorel‡

November 22, 2018

**Abstract.** We investigate an algorithm of gradient type with a backward inertial step in connection with the minimization of a nonconvex differentiable function. We show that the generated sequences converge to a critical point of the objective function, if a regularization of the objective function satisfies the Kurdyka-Łojasiewicz property. Further, we provide convergence rates for the generated sequences and the objective function values formulated in terms of the Łojasiewicz exponent. Finally, some numerical experiments are presented in order to compare our numerical scheme with some algorithms well known in the literature.

**Key Words.** inertial algorithm, nonconvex optimization, Kurdyka-Łojasiewicz inequality, convergence rate

**AMS subject classification.** 90C26, 90C30, 65K10

## 1 Introduction and Preliminaries

Gradient-type algorithms have a long history, going back at least to Cauchy (1847), and also a wealth of applications. Solving linear systems, Cauchy’s original motivation, is maybe the most obvious application, but many of today’s hot topics in machine learning or image processing also deal with optimization problems from an algorithmic perspective and rely on gradient-type algorithms.

The original gradient descent algorithm

$$x_{n+1} = x_n - s\nabla g(x_n),$$

which is precisely an explicit Euler method applied to the gradient flow

$$\dot{x}(t) = -\nabla g(x(t)),$$

does not achieve very good convergence rates and much research has been dedicated to accelerating convergence.

Based on the analogy to mechanical systems, e.g. to the movement, with friction, of a heavy ball in a potential well defined by the objective function  $g$ , Polyak [20] was able to provide the seminal idea

---

\*Tiberiu Popoviciu Institute of Numerical Analysis, Romanian Academy, Cluj-Napoca, Romania and Department of Mathematics, Babes-Bolyai University, Cluj-Napoca, Romania, e-mail: cristian.alecsa@math.ubbcluj.ro. This work was supported by a grant of Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2016-0266, within PNCDI III.

†Technical University of Cluj-Napoca, Department of Mathematics, Str. Memorandumului nr. 28, 400114 Cluj-Napoca, Romania, e-mail: laszlosziszi@yahoo.com. This work was supported by a grant of Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2016-0266, within PNCDI III.

‡Technical University of Cluj-Napoca, Department of Mathematics, Str. Memorandumului nr. 28, 400114 Cluj-Napoca, Romania, e-mail: oviorel@mail.utcluj.ro. This work was supported by a grant of Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2016-0266, within PNCDI III.

for achieving acceleration namely the addition of inertial (momentum) terms to the gradient algorithm. Probably the most acclaimed inertial algorithm is Nesterov's accelerated gradient method, which in its particular form

$$\begin{cases} x_{n+1} = y_n - s\nabla g(y_n), \\ y_n = x_n + \frac{n}{n+3}(x_n - x_{n-1}), \end{cases}$$

for a convex  $g$  with Lipschitz continuous gradient, exhibits an improved convergence rate of  $\mathcal{O}(1/n^2)$  and which, as highlighted by Su, Boyd and Candés [22], can be seen as the discrete counterpart of the second order differential equation

$$\ddot{x}(t) + \frac{3}{t}\dot{x}(t) - \nabla g(x(t)) = 0.$$

In this paper, we deal with the optimization problem

$$\inf_{x \in \mathbb{R}^m} g(x), \quad (\text{P})$$

where  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  is a Fréchet differentiable function with  $L_g$ -Lipschitz continuous gradient, i.e. there exists  $L_g \geq 0$  such that  $\|\nabla g(x) - \nabla g(y)\| \leq L_g\|x - y\|$  for all  $x, y \in \mathbb{R}^m$ , and we associate to (P) the following inertial algorithm of gradient type.

Consider the starting points  $x_0 = y_0 \in \mathbb{R}^m$ , and for every  $n \in \mathbb{N}$  let

$$\begin{cases} x_{n+1} = y_n - \beta_n \nabla g(y_n), \\ y_n = x_n + \alpha_n(x_n - x_{n-1}), \end{cases} \quad (1)$$

where we assume that

$$\lim_{n \rightarrow +\infty} \alpha_n = \alpha \in \left( \frac{-10 + \sqrt{68}}{8}, 0 \right), \quad \lim_{n \rightarrow +\infty} \beta_n = \beta \text{ and } 0 < \beta < \frac{4\alpha^2 + 10\alpha + 2}{L_g(2\alpha + 1)^2}.$$

Observe that the inertial parameter  $\alpha_n$  becomes negative after a number of iteration and this can be viewed as we take a backward inertial step in our algorithm. We emphasize that the analysis of the proposed algorithm is intimately related to the properties of the following regularization of the objective function  $g$  (see also [10, 11, 12, 15, 16]), that is  $H : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ , defined by :

$$H(x, y) = g(x) + M\|y - x\|^2, \quad M > 0. \quad (2)$$

In the remaining of this section we introduce the necessary apparatus of notions and results that we will need in our forthcoming analysis.

For a differentiable function  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  we denote by  $\text{crit}(f) = \{x \in \mathbb{R}^m : \nabla f(x) = 0\}$  the set of critical points of  $f$ . The following so called descent lemma, see [19], will play an essential role in our forthcoming analysis.

**Lemma 1** *Let  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  be Fréchet differentiable with  $L$  Lipschitz continuous gradient. Then*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2, \quad \forall x, y \in \mathbb{R}^m.$$

Furthermore, the set of cluster points of a given sequence  $(x_n)_{n \in \mathbb{N}}$  will be denoted by  $\omega((x_n)_{n \in \mathbb{N}})$ . At the same time, the distance function to a set, is defined for  $A \subseteq \mathbb{R}^m$  as

$$\text{dist}(x, A) = \inf_{y \in A} \|x - y\| \text{ for all } x \in \mathbb{R}^m.$$

Our convergence result relies on the concept of a KL function. For  $\eta \in (0, +\infty]$ , we denote by  $\Theta_\eta$  the class of concave and continuous functions  $\varphi : [0, \eta) \rightarrow [0, +\infty)$  such that  $\varphi(0) = 0$ ,  $\varphi$  is continuously differentiable on  $(0, \eta)$ , continuous at 0 and  $\varphi'(s) > 0$  for all  $s \in (0, \eta)$ .

**Definition 1** (*Kurdyka-Lojasiewicz property*) Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a differentiable function. We say that  $f$  satisfies the *Kurdyka-Lojasiewicz (KL) property* at  $\bar{x} \in \mathbb{R}^n$  if there exists  $\eta \in (0, +\infty]$ , a neighborhood  $U$  of  $\bar{x}$  and a function  $\varphi \in \Theta_\eta$  such that for all  $x$  in the intersection

$$U \cap \{x \in \mathbb{R}^m : f(\bar{x}) < f(x) < f(\bar{x}) + \eta\}$$

the following inequality holds

$$\varphi'(f(x) - f(\bar{x}))\|\nabla f(x)\| \geq 1.$$

If  $f$  satisfies the KL property at each point in  $\mathbb{R}^m$ , then  $f$  is called a *KL function*.

The function  $\varphi$  is called a desingularizing function (see for instance [5]). The origins of this notion go back to the pioneering work of Lojasiewicz [17], where it is proved that for a real-analytic function  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  and a critical point  $\bar{x} \in \mathbb{R}^m$  (that is  $\nabla f(\bar{x}) = 0$ ), there exists  $\theta \in [1/2, 1)$  such that the function  $|f - f(\bar{x})|^\theta \|\nabla f\|^{-1}$  is bounded around  $\bar{x}$ . This corresponds to the situation when  $\varphi(s) = C(1 - \theta)^{-1} s^{1-\theta}$ , where  $C > 0$  is a given constant, and leads to the following definition.

**Definition 2 (for which we refer to [17, 7, 1].)** A differentiable function  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  has the *Lojasiewicz property with exponent  $\theta \in (0, 1)$*  at  $\bar{x} \in \text{crit}(f)$  if there exist  $K, \epsilon > 0$  such that

$$|f(x) - f(\bar{x})|^\theta \leq K \|\nabla f(x)\|, \quad (3)$$

for every  $x \in \mathbb{R}^m$ , with  $\|x - \bar{x}\| < \epsilon$ .

The result of Lojasiewicz allows the interpretation of the KL property as a re-parametrization of the function values in order to avoid flatness around the critical points. Kurdyka [14] extended this property to differentiable functions definable in an o-minimal structure. Further extensions to the nonsmooth setting can be found in [7, 2, 8, 9].

To the class of KL functions belong semi-algebraic, real sub-analytic, semiconvex, uniformly convex and convex functions satisfying a growth condition. We refer the reader to [7, 2, 9, 6, 8, 3, 1] and the references therein for more details regarding all the classes mentioned above and illustrating examples.

Finally, an important role in our convergence analysis will be played by the following uniformized KL property given in [6, Lemma 6].

**Lemma 2** Let  $\Omega \subseteq \mathbb{R}^m$  be a compact set and let  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  be a differentiable function. Assume that  $f$  is constant on  $\Omega$  and  $f$  satisfies the KL property at each point of  $\Omega$ . Then there exist  $\varepsilon, \eta > 0$  and  $\varphi \in \Theta_\eta$  such that for all  $\bar{x} \in \Omega$  and for all  $x$  in the intersection

$$\{x \in \mathbb{R}^m : \text{dist}(x, \Omega) < \varepsilon\} \cap \{x \in \mathbb{R}^m : f(\bar{x}) < f(x) < f(\bar{x}) + \eta\} \quad (4)$$

the following inequality holds

$$\varphi'(f(x) - f(\bar{x}))\|\nabla f(x)\| \geq 1. \quad (5)$$

The outline of the paper is the following. In Section 2 we give a sufficient condition that ensures the decrease property of the regularization  $H$  in the iterates, and which also ensures that the iterates gap belongs to  $l^2$ . Further, using these results, we show that the set of cluster points of the iterates is included in the set of critical points of the objective function. Finally, by means of the the KL property of  $H$  we obtain that the iterates gap belongs to  $l^1$ . This implies the convergence of the iterates, see also [3, 6, 11, 15]. In Section 3, we obtain several convergence rates both for the sequences  $(x_n)_{n \in \mathbb{N}}$ ,  $(y_n)_{n \in \mathbb{N}}$  generated by the numerical scheme (1), as well as for the function values  $g(x_n)$ ,  $g(y_n)$  in the terms of the Lojasiewicz exponent of  $g$  and  $H$ , respectively, see also [13, 15]. Finally, in Section 4, we present some numerical experiment that shows that our algorithm, in many cases, has better properties than the algorithms used in the literature.

## 2 Convergence results

We start to investigate the convergence of the proposed algorithm by showing that  $H$  is decreasing along certain sequences built upon the iterates generated by (1).

**Theorem 3** *Assume that  $g$  is bounded from below and consider the sequences  $(x_n)_{n \in \mathbb{N}}$ ,  $(y_n)_{n \in \mathbb{N}}$  generated by the numerical scheme (1) together with*

$$u_n = \sqrt{\frac{\delta_n}{M}} \cdot (x_n - x_{n-1}) + y_n, \quad (6)$$

where  $\delta_n$  will be explicitly specified later.

Then, there exists  $N \in \mathbb{N}$  such that

(i) *The sequence  $(H(y_n, u_n))_{n \geq N} = (g(y_n) + \delta_n \|x_n - x_{n-1}\|^2)_{n \geq N}$  is decreasing and  $\delta_n > 0$  for all  $n \geq N$ .*

(ii) *The sequence  $(H(y_n, u_n))_{n \geq N}$  is convergent;*

(iii)  $\sum_{n \geq 1} \|x_n - x_{n-1}\|^2 < +\infty$ .

**Remark 4** An interesting fact is that for the sequence  $(H(y_n, u_n))_{n \geq N}$  to be decreasing one does not need the boundedness of the objective function  $g$ , but only its regularity, as can be seen below, in the proof of Theorem 3. The energy decay is thus a structural property of the algorithm (1).

**Proof.** By applying the descent Lemma 1 to  $g(y_{n+1})$  we have

$$g(y_{n+1}) \leq g(y_n) + \langle \nabla g(y_n), y_{n+1} - y_n \rangle + \frac{Lg}{2} \|y_{n+1} - y_n\|^2.$$

However, after rewriting the first equation in (1) as  $\nabla g(y_n) = \frac{1}{\beta_n}(y_n - x_{n+1})$  and taking the inner product with  $y_{n+1} - y_n$  to obtain

$$\langle \nabla g(y_n), y_{n+1} - y_n \rangle = \frac{1}{\beta_n} \langle y_n - x_{n+1}, y_{n+1} - y_n \rangle,$$

the descent inequality becomes

$$g(y_{n+1}) - \frac{Lg}{2} \|y_{n+1} - y_n\|^2 \leq g(y_n) + \frac{1}{\beta_n} \langle y_n - x_{n+1}, y_{n+1} - y_n \rangle. \quad (7)$$

Further,

$$\begin{aligned} \langle y_n - x_{n+1}, y_{n+1} - y_n \rangle &= -\|y_{n+1} - y_n\|^2 + \langle y_{n+1} - x_{n+1}, y_{n+1} - y_n \rangle, \\ y_{n+1} - y_n &= (1 + \alpha_{n+1})(x_{n+1} - x_n) - \alpha_n(x_n - x_{n-1}) \end{aligned}$$

and

$$y_{n+1} - x_{n+1} = \alpha_{n+1}((x_{n+1} - x_n)),$$

hence

$$g(y_{n+1}) + \left( \frac{1}{\beta_n} - \frac{Lg}{2} \right) \|y_{n+1} - y_n\|^2 \leq g(y_n) + \frac{\alpha_{n+1}}{\beta_n} \langle x_{n+1} - x_n, y_{n+1} - y_n \rangle. \quad (8)$$

Thus we have

$$\|y_{n+1} - y_n\|^2 = \|(1 + \alpha_{n+1})(x_{n+1} - x_n) - \alpha_n(x_n - x_{n-1})\|^2 =$$

$$(1 + \alpha_{n+1})^2 \|x_{n+1} - x_n\|^2 + \alpha_n^2 \|x_n - x_{n-1}\|^2 - 2\alpha_n(1 + \alpha_{n+1}) \langle x_{n+1} - x_n, x_n - x_{n-1} \rangle,$$

and

$$\begin{aligned} \langle x_{n+1} - x_n, y_{n+1} - y_n \rangle &= \langle x_{n+1} - x_n, (1 + \alpha_{n+1})(x_{n+1} - x_n) - \alpha_n(x_n - x_{n-1}) \rangle = \\ &= (1 + \alpha_{n+1}) \|x_{n+1} - x_n\|^2 - \alpha_n \langle x_{n+1} - x_n, x_n - x_{n-1} \rangle. \end{aligned}$$

Replacing the above equalities in (8) gives

$$\begin{aligned} g(y_{n+1}) + \frac{(2 - \beta_n L_g)(1 + \alpha_{n+1})^2 - 2\alpha_{n+1}(1 + \alpha_{n+1})}{2\beta_n} \|x_{n+1} - x_n\|^2 \leq \\ g(y_n) - \frac{(2 - \beta_n L_g)\alpha_n^2}{2\beta_n} \|x_n - x_{n-1}\|^2 + \\ \frac{(2 - \beta_n L_g)\alpha_n(1 + \alpha_{n+1}) - \alpha_n\alpha_{n+1}}{\beta_n} \langle x_{n+1} - x_n, x_n - x_{n-1} \rangle. \end{aligned}$$

The above inequality motivates the introduction of the following notations

$$\begin{aligned} A_n &= \frac{(2 - \beta_n L_g)(1 + \alpha_{n+1})^2 - 2\alpha_{n+1}(1 + \alpha_{n+1})}{2\beta_n}, \\ B_n &= \frac{(2 - \beta_n L_g)\alpha_n^2}{2\beta_n}, \\ C_n &= \frac{(2 - \beta_n L_g)\alpha_n(1 + \alpha_{n+1}) - \alpha_n\alpha_{n+1}}{2\beta_n} \end{aligned}$$

and

$$\delta_n := A_{n-1} + C_{n-1} \quad \text{and} \quad \Delta_n = A_{n-1} + B_n + C_{n-1} + C_n \quad (9)$$

for all  $n \in \mathbb{N}$ ,  $n \geq 1$ .

In terms of these notations, after using the equality

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

we can write

$$-C_n \|x_{n+1} - x_{n-1}\|^2 + g(y_{n+1}) + (A_n + C_n) \|x_{n+1} - x_n\|^2 \leq g(y_n) + (-C_n - B_n) \|x_n - x_{n-1}\|^2. \quad (10)$$

Consequently, we have

$$-C_n \|x_{n+1} - x_{n-1}\|^2 + \Delta_n \|x_n - x_{n-1}\|^2 \leq (g(y_n) + \delta_n \|x_n - x_{n-1}\|^2) - (g(y_{n+1}) + \delta_{n+1} \|x_{n+1} - x_n\|^2). \quad (11)$$

Now, since  $\alpha_n \rightarrow \alpha$ ,  $\beta_n \rightarrow \beta$  as  $n \rightarrow +\infty$  and  $\alpha \in \left(\frac{-10 + \sqrt{68}}{8}, 0\right)$ ,  $0 < \beta < \frac{4\alpha^2 + 10\alpha + 2}{L_g(2\alpha + 1)^2}$  we have

$$\begin{aligned} \lim_{n \rightarrow +\infty} A_n &= \frac{(2 - \beta L_g)(\alpha + 1)^2 + 2\alpha - 2\alpha^2}{2\beta}, \\ \lim_{n \rightarrow +\infty} B_n &= \frac{(2 - \beta L_g)\alpha^2}{2\beta}, \\ \lim_{n \rightarrow +\infty} C_n &= \frac{(2 - \beta L_g)\alpha(1 + \alpha) - \alpha^2}{2\beta} < 0, \\ \lim_{n \rightarrow +\infty} \Delta_n &= \frac{(2 - \beta L_g)(2\alpha + 1)^2 + 2\alpha - 4\alpha^2}{2\beta} > 0, \\ \lim_{n \rightarrow +\infty} \delta_n &= \frac{(2 - \beta L_g)(2\alpha^2 + 3\alpha + 1) + 2\alpha - 3\alpha^2}{2\beta} > 0. \end{aligned}$$

Hence, there exists  $N \in \mathbb{N}$  and  $C > 0$ ,  $D > 0$  such that for all  $n \geq N$  one has

$$C_n \leq -C, \Delta_n \geq D \text{ and } \delta_n > 0$$

which, in the view of (11), shows (i), that is, the sequence  $g(y_n) + \delta_n \|x_n - x_{n-1}\|^2$  is decreasing for  $n \geq N$ .

By using (11) again, we obtain

$$0 < C \|x_{n+1} - x_{n-1}\|^2 + D \|x_n - x_{n-1}\|^2 \leq (g(y_n) + \delta_n \|x_n - x_{n-1}\|^2) - (g(y_{n+1}) + \delta_{n+1} \|x_{n+1} - x_n\|^2),$$

for all  $n \geq N$ , or, more convenient, that

$$0 < D \|x_n - x_{n-1}\|^2 \leq (g(y_n) + \delta_n \|x_n - x_{n-1}\|^2) - (g(y_{n+1}) + \delta_{n+1} \|x_{n+1} - x_n\|^2), \quad (12)$$

for all  $n \geq N$ . Let  $r > N$ . Summing up the latter relations gives

$$D \sum_{n=N}^r \|x_n - x_{n-1}\|^2 \leq (g(y_N) + \delta_N \|x_N - x_{N-1}\|^2) - (g(y_{r+1}) + \delta_{r+1} \|x_{r+1} - x_r\|^2)$$

which leads to

$$g(y_{r+1}) + D \sum_{n=N}^r \|x_n - x_{n-1}\|^2 \leq g(y_N) + \delta_N \|x_N - x_{N-1}\|^2. \quad (13)$$

Now, taking into account that  $g$  is bounded from below, after letting  $r \rightarrow +\infty$  we obtain

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

which proves (iii).

This also shows that

$$\lim_{n \rightarrow +\infty} \|x_n - x_{n-1}\|^2 = 0,$$

hence

$$\lim_{n \rightarrow +\infty} \delta_n \|x_n - x_{n-1}\|^2 = 0.$$

But then, using again the fact that  $g$  is bounded from below we have that the sequence  $g(y_n) + \delta_n \|x_n - x_{n-1}\|^2$  is bounded from below and also decreasing (see (i)) for  $n \geq N$ , hence there exists

$$\lim_{n \rightarrow +\infty} g(y_n) + \delta_n \|x_n - x_{n-1}\|^2 \in \mathbb{R}.$$

■

**Remark 5** Observe that conclusion (iii) in the hypotheses of Theorem 3 assures that the sequence  $(x_n - x_{n-1})_{n \in \mathbb{N}} \in l^2$ , in particular that

$$\lim_{n \rightarrow +\infty} (x_n - x_{n-1}) = 0. \quad (14)$$

**Lemma 6** *In the framework of the optimization problem (P) assume that the objective function  $g$  is bounded from below and consider the sequences  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  generated by the numerical algorithm (1) and let  $(u_n)_{n \in \mathbb{N}}$  defined by (6). Then, the following statements are valid :*

- (i) *the sets of cluster points of  $(x_n)_{n \in \mathbb{N}}$ ,  $(y_n)_{n \in \mathbb{N}}$  and  $(u_n)_{n \in \mathbb{N}}$  coincide and are contained in the set of critical points of  $g$ , i.e.,*

$$\omega((x_n)_{n \in \mathbb{N}}) = \omega((y_n)_{n \in \mathbb{N}}) = \omega((u_n)_{n \in \mathbb{N}}) \subseteq \text{crit}(g);$$

$$(ii) \quad \omega((y_n, u_n)_{n \in \mathbb{N}}) \subseteq \text{crit}(H) = \{(x, x) \in \mathbb{R}^m \times \mathbb{R}^m : x \in \text{crit}(g)\} .$$

**Proof.** (i) We start by proving  $\omega((x_n)_{n \in \mathbb{N}}) \subseteq \omega((u_n)_{n \in \mathbb{N}})$  and  $\omega((x_n)_{n \in \mathbb{N}}) \subseteq \omega((y_n)_{n \in \mathbb{N}})$ . Bearing in mind that  $\lim_{n \rightarrow \infty} (x_n - x_{n-1}) = 0$  and that the sequences  $(\delta_n)_{n \in \mathbb{N}}$ ,  $(\alpha_n)_{n \in \mathbb{N}}$  and  $(\beta_n)_{n \in \mathbb{N}}$  are convergent, the conclusion is quite straightforward. Indeed, if  $\bar{x} \in \omega((x_n)_{n \in \mathbb{N}})$  and  $(x_{n_k})_{k \in \mathbb{N}}$  is a subsequence such that  $\lim_{k \rightarrow \infty} x_{n_k} = \bar{x}$  then

$$\lim_{k \rightarrow \infty} y_{n_k} = \lim_{k \rightarrow \infty} x_{n_k} + \lim_{k \rightarrow \infty} \alpha_{n_k} \cdot \lim_{k \rightarrow \infty} (x_{n_k} - x_{n_k-1})$$

and

$$\lim_{k \rightarrow \infty} u_{n_k} = \lim_{k \rightarrow \infty} \delta_{n_k} \cdot \lim_{k \rightarrow \infty} (x_{n_k} - x_{n_k-1}) + \lim_{k \rightarrow \infty} y_{n_k}$$

imply that the sequences  $(x_{n_k})_{k \in \mathbb{N}}$ ,  $(y_{n_k})_{k \in \mathbb{N}}$  and  $(u_{n_k})_{k \in \mathbb{N}}$  all converge to the same element  $\bar{x} \in \mathbb{R}^m$ .

In order to prove that  $\omega((x_n)_{n \in \mathbb{N}}) \subseteq \text{crit}(g)$  we use the fact that  $\nabla g$  is a continuous operator. So, passing to the limit in  $\nabla g(y_{n_k}) = \frac{1}{\beta_{n_k}} \cdot (y_{n_k} - x_{n_k+1})$  gives

$$\begin{aligned} \nabla g(\bar{x}) &= \lim_{k \rightarrow \infty} \nabla g(y_{n_k}) \\ &= \frac{1}{\lim_{k \rightarrow \infty} \beta_{n_k}} \cdot \lim_{k \rightarrow \infty} (y_{n_k} - x_{n_k+1}) \end{aligned}$$

and finally, as  $y_{n_k} - x_{n_k+1} = (x_{n_k} - x_{n_k+1}) + \alpha_{n_k} \cdot (x_{n_k} - x_{n_k-1})$ , to

$$\nabla g(\bar{x}) = 0.$$

The reverse inclusions follow in a very similar manner from the definitions of  $u_n$  and  $y_n$ .

For proving the statement (ii), we rely on a direct computation yielding

$$\nabla H(x, y) = (\nabla g(x) + 2M(x - y), 2M(y - x)), \quad (15)$$

which, in turn, gives

$$\text{crit}(H) = \{(\bar{x}, \bar{x}) \in \mathbb{R}^m \times \mathbb{R}^m : \bar{x} \in \text{crit}(g)\}$$

and allows us to apply (i) to obtain the desired conclusion. ■

Some direct consequences of Theorem 3 (ii) and Lemma 6 are the following.

**Fact 7** *In the setting of Lemma 6, let  $(\bar{x}, \bar{x}) \in \omega((y_n, u_n)_{n \in \mathbb{N}})$ . It follows that  $\bar{x} \in \text{crit}(g)$  and*

$$\lim_{n \rightarrow \infty} H(y_n, u_n) = H(\bar{x}, \bar{x}).$$

Consequently,

$$H \text{ is finite and constant on the set } \omega((y_n, u_n)_{n \in \mathbb{N}}).$$

The arguments behind the proofs of the following facts are the same as those in Lemma 4 from [15].

**Fact 8** *If the assumptions from Lemma 6 hold true and if the sequence  $(x_n)_{n \in \mathbb{N}}$  is bounded, then the following conclusions hold up :*

- (i)  $\omega((y_n, u_n)_{n \in \mathbb{N}})$  is nonempty and compact ,
- (ii)  $\lim_{n \rightarrow +\infty} \text{dist}((y_n, u_n), \omega((y_n, u_n)_{n \in \mathbb{N}})) = 0$  .

**Remark 9** We emphasize that if  $g$  is coercive, that is  $\lim_{\|x\| \rightarrow +\infty} g(x) = +\infty$ , then  $g$  is bounded from below and  $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ , the sequences generated by (1), are bounded.

Indeed, notice that  $g$  is bounded from below, being a continuous and coercive function (see [21]). Note that according to Theorem 3 the sequence  $D \sum_{n=N}^r \|x_n - x_{n-1}\|^2$  is convergent hence is bounded. Consequently, from (13) it follows that  $y_r$  is contained for every  $r > N$ , ( $N$  is defined in the hypothesis of Theorem 3), in a lower level set of  $g$ , which is bounded. Since  $(y_n)_{n \in \mathbb{N}}$  is bounded, taking into account (14), it follows that  $(x_n)_{n \in \mathbb{N}}$  is also bounded.

Now, based on the conclusions of Lemma 6, we present a result which will be crucial latter on. For our next result,  $\|\cdot\|_1$  will denote the 1-norm and  $\|\cdot\|_2$  will represent the 2-norm on the linear space  $\mathbb{R}^m \times \mathbb{R}^m$ .

**Lemma 10** *Let  $H, (x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$  and  $(u_n)_{n \in \mathbb{N}}$  be as in all the previous results, with the mapping  $g$  bounded from below. Then, the following gradient inequalities hold true*

$$\|\nabla H(y_n, u_n)\|_2 \leq \|\nabla H(y_n, u_n)\|_1 \leq \frac{1}{\beta_n} \cdot \|x_{n+1} - x_n\| + \left[ \left( \frac{\alpha_n}{\beta_n} \right) + 4M \sqrt{\frac{\delta_n}{M}} \right] \cdot \|x_n - x_{n-1}\| \quad (16)$$

and

$$\|\nabla H(y_n, u_n)\|_2^2 \leq \frac{2}{\beta_n^2} \cdot \|x_{n+1} - x_n\|^2 + 2 \left[ \left( \frac{\alpha_n}{\beta_n} - 2M \sqrt{\frac{\delta_n}{M}} \right)^2 + 2\delta_n M \right] \cdot \|x_n - x_{n-1}\|^2. \quad (17)$$

**Proof.** First of all note that from our numerical scheme (1) we have  $\nabla g(y_n) = \frac{1}{\beta_n}((x_n - x_{n+1}) + \alpha_n(x_n - x_{n-1}))$ . In terms of the  $\|\cdot\|_1$  on  $\mathbb{R}^m \times \mathbb{R}^m$ , we have

$$\begin{aligned} \|\nabla H(y_n, u_n)\|_1 &= \|(\nabla g(y_n) + 2M(y_n - u_n), 2M(u_n - y_n))\|_1 \\ &= \|\nabla g(y_n) + 2M(y_n - u_n)\| + 2M\|u_n - y_n\| \\ &\leq \frac{1}{\beta_n} \|x_{n+1} - x_n\| + \frac{\alpha_n}{\beta_n} \|x_n - x_{n-1}\| + 4M \sqrt{\frac{\delta_n}{M}} \|x_n - x_{n-1}\|, \end{aligned}$$

which proves the desired inequality.

Now, with respect to the Euclidean norm, similar arguments yield

$$\begin{aligned} \|\nabla H(y_n, u_n)\|_2^2 &= \|\nabla g(y_n) + 2M(y_n - u_n)\|^2 + \|2M(u_n - y_n)\|^2 \\ &= \left\| \nabla g(y_n) - 2M \sqrt{\frac{\delta_n}{M}} (x_n - x_{n-1}) \right\|^2 + 4M^2 \left( \sqrt{\frac{\delta_n}{M}} \right)^2 \cdot \|x_n - x_{n-1}\|^2 \\ &\leq \frac{2}{\beta_n^2} \|x_{n+1} - x_n\|^2 + 2 \left\| \left( \frac{\alpha_n}{\beta_n} - 2M \sqrt{\frac{\delta_n}{M}} \right) \cdot (x_n - x_{n-1}) \right\|^2 + 4M\delta_n \cdot \|x_n - x_{n-1}\|^2, \end{aligned}$$

completing the proof. ■

Our main result concerning the convergence of the sequence  $(x_n)_{n \in \mathbb{N}}$  generated by the algorithm (1) to a critical point of the objective function  $g$  is the following.

**Theorem 11** *Consider the sequences  $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$  generated by the algorithm (1) and let the objective function  $g$  be bounded from below. If the sequence  $(x_n)_{n \in \mathbb{N}}$  is bounded and  $H$  is a KL function, then*

$$\sum_{n=1}^{\infty} \|x_n - x_{n-1}\| < +\infty \quad (18)$$

and there exists an element  $\bar{x} \in \text{crit}(g)$  such that  $\lim_{n \rightarrow +\infty} x_n = \bar{x}$ .



**Proof.** Consider  $(\bar{x}, \bar{x})$  from the set  $\omega((y_n, u_n)_{n \in \mathbb{N}})$  under the assumptions of Lemma (6). It follows that  $\bar{x} \in \text{crit}(g)$ . Also, using Fact (7), we get that  $\lim_{n \rightarrow \infty} H(y_n, u_n) = H(\bar{x}, \bar{x})$ . Furthermore, we consider two cases:

**I.** By using  $N$  from Theorem (3), assume that there exists  $\bar{n} \geq N$ , with  $\bar{n} \in \mathbb{N}$ , such that  $H(y_{\bar{n}}, u_{\bar{n}}) = H(\bar{x}, \bar{x})$ . Then, since  $(H(y_n, u_n))_{n \geq \bar{n}}$  is a decreasing sequence, it follows that

$$H(y_n, u_n) = H(\bar{x}, \bar{x}), \text{ for every } n \geq \bar{n},$$

Now, using (12) we get that for every  $n \geq \bar{n}$  we have the following inequality:

$$0 \leq D \|x_n - x_{n-1}\|^2 \leq H(y_n, u_n) - H(y_{n+1}, u_{n+1}) = H(\bar{x}, \bar{x}) - H(\bar{x}, \bar{x}) = 0.$$

So, the sequence  $(x_n)_{n \geq \bar{n}}$  is constant and the conclusion holds true.

**II.** Now, we deal with the case when  $H(y_n, u_n) > H(\bar{x}, \bar{x})$ , for every  $n \geq N$ .

So, consider the set  $\Omega := \omega((y_n, u_n)_{n \in \mathbb{N}})$ . From Fact (8) we have that the set  $\Omega$  is nonempty and compact. Also, Fact (7) assures that mapping  $H$  is constant on  $\Omega$ . From the hypotheses of the theorem we have that  $H$  is a  $KL$  function. So, according to Lemma 2, there exists  $\varepsilon > 0$ ,  $\eta > 0$  and a function  $\varphi \in \Theta_\eta$ , such that for all the points  $(z, w)$  from the set

$$\{(z, w) \in \mathbb{R}^m \times \mathbb{R}^m : \text{dist}((z, w), \Omega) < \varepsilon\} \cap \{(z, w) \in \mathbb{R}^m \times \mathbb{R}^m : H(\bar{x}, \bar{x}) < H(z, w) < \eta + H(\bar{x}, \bar{x})\}$$

one has that

$$\varphi'(H(z, w) - H(\bar{x}, \bar{x})) \cdot \|\nabla H(z, w)\| \geq 1.$$

On the other hand, using Fact (8), we obtain that  $\lim_{n \rightarrow +\infty} \text{dist}((y_n, u_n), \Omega) = 0$ . This means that there exists an index  $n_1 \in \mathbb{N}$ , for which it is valid

$$\text{dist}((y_n, u_n), \Omega) < \varepsilon, \text{ for all } n \geq n_1.$$

Because

$$\lim_{n \rightarrow +\infty} H(y_n, u_n) = H(\bar{x}, \bar{x})$$

and since

$$H(y_n, u_n) > H(\bar{x}, \bar{x}), \text{ for all } n \geq N$$

then there exists another index  $n_2 \geq N$ , such that

$$H(\bar{x}, \bar{x}) < H(y_n, u_n) < H(\bar{x}, \bar{x}) + \eta, \text{ for every } n \geq n_2.$$

Taking  $\bar{n} := \max(n_1, n_2)$  we get that for each  $n \geq \bar{n}$  it follows

$$\varphi'(H(y_n, u_n) - H(\bar{x}, \bar{x})) \cdot \|\nabla H(y_n, u_n)\| \geq 1.$$

Since the function  $\varphi$  is concave, we have

$$\begin{aligned} & \varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x})) \geq \\ & \varphi'(H(y_n, u_n) - H(\bar{x}, \bar{x})) \cdot (H(y_n, u_n) - H(y_{n+1}, u_{n+1})). \end{aligned}$$

Thus, the following relation takes place for each  $n \geq \bar{n}$ :

$$\varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x})) \geq \frac{H(y_n, u_n) - H(y_{n+1}, u_{n+1})}{\|\nabla H(y_n, u_n)\|}.$$

On one hand, combining the inequality (12) and (16), it follows that for every  $n \geq \bar{n}$

$$\begin{aligned} & \varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x})) \geq \\ & \frac{D\|x_n - x_{n-1}\|^2}{\frac{1}{\beta_n}\|x_n - x_{n+1}\| + \left[ \frac{\alpha_n}{\beta_n} + 4M\sqrt{\frac{\delta_n}{M}} \right] \cdot \|x_n - x_{n-1}\|}. \end{aligned} \quad (19)$$

On the other hand, we know that the sequences  $(\alpha_n)_{n \in \mathbb{N}}$ ,  $(\beta_n)_{n \in \mathbb{N}}$  and  $(\delta_n)_{n \in \mathbb{N}}$  are convergent, and  $\lim_{n \rightarrow +\infty} \beta_n = \beta > 0$ , hence  $\left( \frac{1}{\beta_n} \right)_{n \in \mathbb{N}}$  and  $\left( \frac{\alpha_n}{\beta_n} + 4M\sqrt{\frac{\delta_n}{M}} \right)_{n \in \mathbb{N}}$  are bounded. This shows that there exists  $\bar{N} \in \mathbb{N}$ ,  $\bar{N} \geq \bar{n}$  and there exists  $M' > 0$ , such that

$$\sup_{n \geq \bar{N}} \left\{ \frac{1}{\beta_n}, \frac{\alpha_n}{\beta_n} + 4M\sqrt{\frac{\delta_n}{M}} \right\} \leq M'.$$

Thus, the inequality (19) becomes

$$\begin{aligned} & \varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x})) \geq \\ & \frac{D\|x_n - x_{n-1}\|^2}{M'(\|x_n - x_{n+1}\| + \|x_n - x_{n-1}\|)}, \end{aligned} \quad (20)$$

for every  $n \geq \bar{N}$ . This implies that for each  $n \geq \bar{N}$ , the following inequality holds :

$$\|x_n - x_{n-1}\| \leq \sqrt{\frac{M'}{D} \cdot (\varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x}))) \cdot (\|x_n - x_{n+1}\| + \|x_n - x_{n-1}\|)}.$$

From the well-known arithmetical-geometrical inequality, it follows that

$$\begin{aligned} & \sqrt{\frac{M'}{D} \cdot (\varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x}))) \cdot (\|x_n - x_{n+1}\| + \|x_n - x_{n-1}\|)} \leq \\ & \frac{\|x_{n+1} - x_n\| + \|x_n - x_{n-1}\|}{3} + \frac{3M'}{4D} \cdot (\varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x}))). \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} & \|x_n - x_{n-1}\| \leq \\ & \frac{\|x_{n+1} - x_n\| + \|x_n - x_{n-1}\|}{3} + \frac{3M'}{4D} \cdot (\varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x}))). \end{aligned}$$

Consequently, we have

$$2\|x_n - x_{n-1}\| - \|x_n - x_{n+1}\| \leq \quad (21)$$

$$\frac{9M'}{4D} \cdot (\varphi(H(y_n, u_n) - H(\bar{x}, \bar{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\bar{x}, \bar{x}))),$$

for every  $n \in \mathbb{N}$ , with  $n \geq \bar{N}$ . Now, by summing up the latter inequality from  $\bar{N}$  to  $P \geq \bar{N}$ , we get that

$$\begin{aligned} & \sum_{n=\bar{N}}^P \|x_n - x_{n-1}\| \leq \|x_{P+1} - x_P\| - \|x_{\bar{N}} - x_{\bar{N}-1}\| \\ & + \frac{9M'}{4D} \cdot (\varphi(H(y_{\bar{N}}, u_{\bar{N}}) - H(\bar{x}, \bar{x})) - \varphi(H(y_{P+1}, u_{P+1}) - H(\bar{x}, \bar{x}))). \end{aligned}$$

Now, it is time to use the fact that  $\varphi(0) = 0$ . In this setting, by letting  $P \rightarrow +\infty$  and by using (14) we obtain

$$\sum_{n=\bar{N}}^{\infty} \|x_n - x_{n-1}\| \leq -\|x_{\bar{N}} - x_{\bar{N}-1}\| + \frac{9M'}{4D} \varphi(H(y_{\bar{N}}, u_{\bar{N}}) - H(\bar{x}, \bar{x})) < +\infty.$$

It implies that

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

so the first part of the proof is done.

At the same time, the sequence  $(S_n)_{n \in \mathbb{N}}$ , defined by

$$S_n = \sum_{i=1}^n \|x_i - x_{i-1}\|$$

is Cauchy. Thus, for every  $\varepsilon > 0$ , there exists a positive integer number  $N_\varepsilon$ , such that for each  $n \geq N_\varepsilon$  and for all  $p \in \mathbb{N}$ , one has

$$S_{n+p} - S_n \leq \varepsilon.$$

Furthermore,

$$S_{n+p} - S_n = \sum_{i=n+1}^{n+p} \|x_i - x_{i-1}\| \geq \left\| \sum_{i=n+1}^{n+p} (x_i - x_{i-1}) \right\| = \|x_{n+p} - x_n\|.$$

So, the sequence  $(x_n)_{n \in \mathbb{N}}$  is Cauchy, hence is convergent, i.e. there exists  $x \in \mathbb{R}^m$ , such that

$$\lim_{n \rightarrow +\infty} x_n = x.$$

Thus, by using (i) of Lemma (6), it follows that

$$\{x\} = \omega((x_n)_{n \in \mathbb{N}}) \subseteq \text{crit}(g),$$

which leads to the conclusion of the second part of the present theorem. ■

**Remark 12** Since the class of semi-algebraic functions is closed under addition (see for example [6]) and  $(x, y) \mapsto M\|x - y\|^2$  is semi-algebraic, the conclusion of the previous theorem holds if the condition  $H$  is a KL function is replaced by the assumption that  $g$  is semi-algebraic.

Note that, according to Remark 9, the conclusion of Theorem 11 remains valid if we replace in its hypotheses the conditions that  $g$  is bounded from below and  $(x_n)_{n \in \mathbb{N}}$  is bounded by the condition that  $g$  is coercive.

Note that under the assumptions of Theorem 11 we have  $\lim_{n \rightarrow +\infty} y_n = x$  and

$$\lim_{n \rightarrow +\infty} g(x_n) = \lim_{n \rightarrow +\infty} g(y_n) = g(x).$$

### 3 Convergence rates

In the following theorems we provide convergence rates for the sequence generated by (1), but also for the function values, in terms of the Łojasiewicz exponent of  $H$  (see, also, [1, 7, 15]). Note that the forthcoming results remain valid if one replace in their hypotheses the conditions that  $g$  is bounded from below and  $(x_n)_{n \in \mathbb{N}}$  is bounded by the condition that  $g$  is coercive.

**Theorem 13** *In the settings of problem (P) consider the sequences  $(x_n)_{n \in \mathbb{N}}$ ,  $(y_n)_{n \in \mathbb{N}}$  generated by Algorithm (1). Assume that  $g$  is bounded from below and that  $(x_n)_{n \in \mathbb{N}}$  is bounded, let  $\bar{x} \in \text{crit}(g)$  be such that  $\lim_{n \rightarrow +\infty} x_n = \bar{x}$  and suppose that*

$$H : \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}, H(x, y) = g(x) + M\|x - y\|^2$$

*fulfills the Lojasiewicz property at  $(\bar{x}, \bar{x}) \in \text{crit } H$  with Lojasiewicz exponent  $\theta \in (0, \frac{1}{2}]$ . Then, for every  $p > 0$  there exist  $a_1, a_2, a_3, a_4 > 0$  and  $\bar{k} \in \mathbb{N}$  such that the following statements hold true:*

$$(a_1) \quad g(y_n) - g(\bar{x}) \leq a_1 \frac{1}{n^p} \text{ for every } n \geq \bar{k},$$

$$(a_2) \quad g(x_n) - g(\bar{x}) \leq a_2 \frac{1}{n^p} \text{ for every } n \geq \bar{k},$$

$$(a_3) \quad \|x_n - \bar{x}\| \leq a_3 \frac{1}{n^{\frac{p}{2}}} \text{ for every } n \geq \bar{k},$$

$$(a_4) \quad \|y_n - \bar{x}\| \leq a_4 \frac{1}{n^{\frac{p}{2}}} \text{ for all } n \geq \bar{k}.$$

**Proof.** We start by employing the ideas from the proof of Theorem 3, namely if there exists  $\bar{n} \in \mathbb{N}$ , with  $\bar{n} \geq N$ , for which one has that

$$H(y_{\bar{n}}, u_{\bar{n}}) = H(\bar{x}, \bar{x}),$$

then it follows that the sequence  $(x_n)_{n \geq \bar{n}}$  is constant. This leads to the fact that the sequence  $(y_n)_{n \geq \bar{n}}$  is also constant. Furthermore,

$$H(y_n, u_n) = H(\bar{x}, \bar{x}) \quad \text{for all } n \geq \bar{n}.$$

That is, if the regularized energy is constant after a certain number of iterations, one can see that the conclusion follow in a straightforward way.

Now, we can easily assume that

$$H(y_n, u_n) > H(\bar{x}, \bar{x}) \quad \text{for all } n \geq N.$$

In order to simplify notations we will use

$$H_n := H(y_n, u_n), \quad \bar{H} := H(\bar{x}, \bar{x}) \quad \text{and} \quad \nabla H_n := \nabla H(y_n, u_n).$$

Our analysis aims at deriving a difference inequality for

$$r_n := H_n - \bar{H} = H(y_n, u_n) - H(\bar{x}, \bar{x}) > 0$$

based on three previously established fundamental relations:

a) the energy decay relation (12), for every  $n \geq N$

$$H_n - H_{n+1} \geq D\|x_n - x_{n-1}\|^2,$$

b) the energy-gradient estimate (17), for every  $n \in \mathbb{N}$

$$\|\nabla H_n\|^2 \leq \frac{2}{\beta_n^2} \|x_{n+1} - x_n\|^2 + S_n \|x_n - x_{n-1}\|^2$$

where

$$S_n = 2 \cdot \left[ \left( \frac{\alpha_n}{\beta_n} - 2M \sqrt{\frac{\delta_n}{M}} \right)^2 + 2\delta_n M \right],$$

c) the Lojasiewicz inequality (3), for every  $n \geq \bar{N}_1$

$$(H_n - \bar{H})^{2\theta} \leq K^2 \|\nabla H_n\|^2,$$

where  $\bar{N}_1$  is defined by the Lojasiewicz property of  $H$  at  $(\bar{x}, \bar{x})$  such that for  $\epsilon > 0$  one has

$$(y_n, u_n) - (\bar{x}, \bar{x}) \|\leq \epsilon$$

for all  $n \geq \bar{N}_1$ .

By combining these three inequalities, one reaches

$$(H_n - \bar{H})^{2\theta} \leq \frac{2K^2}{\beta_n^2 D} (H_{n+1} - H_{n+2}) + \frac{K^2 S_n}{D} (H_n - H_{n+1})$$

and we are led to a nonlinear second order difference inequality

$$r_n^{2\theta} \leq \frac{2K^2}{\beta_n^2 D} (r_{n+1} - r_{n+2}) + \frac{K^2 S_n}{D} (r_n - r_{n+1}), \quad (22)$$

for every  $n \geq \bar{N}_1$ .

Using the fact that the positive sequence  $(r_n)_{n \in \mathbb{N}}$  is decreasing and converging to zero we have that  $r_{n+1} \leq r_n$  and that  $r_n < 1$  for  $n$  large enough. These observations lead, in the view  $\theta \in (0, 1/2]$ , to the existence of  $\bar{N}_2 \geq \bar{N}_1$ ,  $\bar{N}_2 \in \mathbb{N}$  such that  $r_n^{2\theta} \geq r_{n+1}^{2\theta} \geq r_{n+1}$  for all  $n \geq \bar{N}_2$  and (22) reduces to the linear,  $\theta$ -independent difference inequality

$$r_n \geq \mu_n r_{n+2} + (1 + \lambda_n - \mu_n) r_{n+1}, \quad (23)$$

for every  $n \geq \bar{N}_2$ , with the shorthand notations

$$\lambda_n := \frac{D}{K^2 S_n} \quad \text{and} \quad \mu_n := \frac{2}{\beta_n^2 S_n}.$$

In order to derive our first convergence rate we proceed as follows: for any  $p > 0$  we consider the diverging sequence  $(\xi_n)_{n \geq \bar{N}_2}$ ,  $\xi_n \rightarrow +\infty$ , defined by

$$\xi_n := \frac{\mu_n n^p}{(n+1)^p - n^p}.$$

Since

$$\lim_{n \rightarrow \infty} (1 + \lambda_n - \mu_n) - \left( \frac{\mu_{n+1}}{\xi_{n+1}} - \frac{\mu_{n-1}}{1 + \frac{\mu_n}{\xi_n}} + 1 \right) = \lim_{n \rightarrow \infty} \lambda_n > 0,$$

there exists a large enough index  $\bar{k} \in \mathbb{N}$ ,  $\bar{k} \geq \bar{N}_2$ , such that for each  $n \geq \bar{k}$ , one has

$$1 + \lambda_n - \mu_n \geq \frac{\mu_{n+1}}{\xi_{n+1}} - \frac{\mu_{n-1}}{1 + \frac{\mu_n}{\xi_n}} + 1 \quad (24)$$

Combining the above inequality with the difference inequality for  $r_n$  yields after some computations

$$r_n + \frac{\mu_{n-1}}{1 + \frac{\mu_n}{\xi_n}} r_{n+1} \geq \left( 1 + \frac{\mu_{n+1}}{\xi_{n+1}} \right) \cdot \left( r_{n+1} + \frac{\mu_n}{1 + \frac{\mu_{n+1}}{\xi_{n+1}}} r_{n+2} \right)$$

for each  $n \geq \bar{k}$ , and after multiplying these inequalities from  $\bar{k}$  to  $n$  and simplifying we have

$$\left( r_{\bar{k}} + r_{\bar{k}+1} \frac{\mu_{\bar{k}-1}}{1 + \frac{\mu_{\bar{k}}}{\xi_{\bar{k}}}} \right) \prod_{i=\bar{k}}^n \frac{1}{1 + \frac{\mu_{i+1}}{\xi_{i+1}}} \geq \left( r_{n+1} + r_{n+2} \frac{\mu_n}{1 + \frac{\mu_{n+1}}{\xi_{n+1}}} \right). \quad (25)$$

Now, using the fact that  $\frac{\mu_{n+1}}{\xi_{n+1}} = \left( \frac{n+2}{n+1} \right)^p - 1$ , which implies

$$\prod_{i=\bar{k}}^n \frac{1}{1 + \frac{\mu_{i+1}}{\xi_{i+1}}} = \prod_{i=\bar{k}}^n \left( \frac{i+1}{i+2} \right)^p = \left( \frac{\bar{k}+1}{n+2} \right)^p,$$

we obtain

$$a_1 \left( \frac{1}{n+2} \right)^p \geq r_{n+1} + r_{n+2} \frac{\mu_n}{1 + \frac{\mu_{n+1}}{\xi_{n+1}}}$$

where  $a_1 = \left( r_{\bar{k}} + r_{\bar{k}+1} \frac{\mu_{\bar{k}-1}}{1 + \frac{\mu_{\bar{k}}}{\xi_{\bar{k}}}} \right) (\bar{k}+1)^p$ . Now, from the definition of  $r_n$

$$a_1 \left( \frac{1}{n} \right)^p \geq a_1 \left( \frac{1}{n+1} \right)^p \geq r_n = g(y_n) - g(\bar{x}) + \delta_n \|x_n - x_{n-1}\|^2,$$

finally yields the desired estimate

$$g(y_n) - g(\bar{x}) \leq r_n \leq a_1 \cdot \frac{1}{n^p}. \quad (26)$$

In order to give an upper bound for the difference  $g(x_n) - g(\bar{x})$ , we consider the following chain of inequalities based upon Lemma 1:

$$\begin{aligned} g(x_n) - g(y_n) &\leq \langle \nabla g(y_n), x_n - y_n \rangle + \frac{Lg}{2} \|x_n - y_n\|^2 \\ &= \left\langle \frac{1}{\beta_n} (y_n - x_{n+1}), -\alpha_n (x_n - x_{n-1}) \right\rangle + \frac{Lg}{2} \|x_n - y_n\|^2 \\ &= \frac{1}{\beta_n} \langle x_{n+1} - x_n, \alpha_n (x_n - x_{n-1}) \rangle - \alpha_n^2 \frac{2 - \beta_n Lg}{2\beta_n} \|x_n - x_{n-1}\|^2. \end{aligned}$$

Here, using the inequality

$$\langle x_{n+1} - x_n, \alpha_n (x_n - x_{n-1}) \rangle \leq \frac{1}{2} \left[ \frac{1}{2 - \beta_n Lg} \|x_{n+1} - x_n\|^2 + (2 - \beta_n Lg) \alpha_n^2 \|x_n - x_{n-1}\|^2 \right],$$

leads, after some simplifications, to

$$g(x_n) - g(y_n) \leq \frac{1}{2\beta_n(2 - \beta_n Lg)} \|x_{n+1} - x_n\|^2.$$

By combining the inequality (12) with the fact that the sequence  $(g(y_n) + \delta_n \|x_n - x_{n-1}\|^2)_{n \geq \bar{k}}$  is decreasing and converges to  $g(\bar{x})$ , one obtains

$$g(x_n) - g(y_n) \leq \frac{1}{2D\beta_n(2 - \beta_n Lg)} r_{n+1}, \quad \text{for all } n \geq \bar{k}. \quad (27)$$

From (26) we have  $r_{n+1} \leq a_1 \frac{1}{(n+1)^p} \leq a_1 \frac{1}{n^p}$ , hence

$$g(x_n) - g(y_n) \leq \frac{1}{2D\beta_n(2 - \beta_n L_g)} a_1 \frac{1}{n^p}, \text{ for all } n \geq \bar{k}.$$

This means that for every  $n > \bar{k}$  one has

$$g(x_n) - g(\bar{x}) = (g(x_n) - g(y_n)) + (g(y_n) - g(\bar{x})) \leq a_1 \left[ \frac{1}{2D\beta_n(2 - \beta_n L_g)} + 1 \right] \frac{1}{n^p}.$$

Since the sequence  $(\beta_n)_{n \in \mathbb{N}}$  is convergent to  $\beta > 0$ , we can choose

$$a_2 = a_1 \sup_{n \geq \bar{k}} \frac{1}{2D\beta_n(2 - \beta_n L_g)} + a_1$$

and we have

$$g(x_n) - g(\bar{x}) \leq a_2 \frac{1}{n^p}, \text{ for every } n \geq \bar{k}. \quad (28)$$

We continue the proof by establishing an estimate for  $\|x_n - \bar{x}\|$ . By the triangle inequality and by summing up (21) from  $n \geq \bar{k}$  to  $P > n$  one has

$$\begin{aligned} \|x_P - x_{n-1}\| &\leq \sum_{k=n}^P \|x_k - x_{k-1}\| \\ &\leq -\|x_n - x_{n-1}\| + \|x_{P+1} - x_P\| + \frac{9M'}{4D} [\varphi(H_n - \bar{H}) - \varphi(H_{P+1} - \bar{H})], \end{aligned}$$

so, letting  $P \rightarrow \infty$  gives

$$\|x_{n-1} - \bar{x}\| \leq \frac{9M'}{4D} \varphi(H_n - \bar{H}).$$

Recall, however, that the desingularizing function is  $\varphi(t) = \frac{K}{1-\theta} t^{1-\theta}$  hence,

$$\|x_{n-1} - \bar{x}\| \leq M_1 r_n^{1-\theta}, \quad (29)$$

with  $M_1 = \frac{9M'K}{4D(1-\theta)}$ . Further, due to  $0 < r_n < 1$  combined with  $r_n^{1-\theta} \leq \sqrt{r_n}$  which holds for  $\theta \in (0, 1/2]$ , we conclude that

$$\|x_n - \bar{x}\| \leq M_1 \sqrt{r_{n+1}} \leq M_1 \sqrt{r_n} \leq a_3 \left( \frac{1}{n} \right)^{p/2}, \text{ for every } n \geq \bar{k}, \quad (30)$$

where  $a_3 := \sqrt{a_1} M_1$ .

Finally, we conclude this part of the proof by deducing an upper bound for  $\|y_n - \bar{x}\|$ . The following inequalities hold true

$$\begin{aligned} \|y_n - \bar{x}\| &= \left\| x_n + \alpha_n(x_n - x_{n-1}) - \bar{x} \right\| \\ &\leq |1 + \alpha_n| \cdot \|x_n - \bar{x}\| + |\alpha_n| \cdot \|x_{n-1} - \bar{x}\| \\ &\leq (1 + |\alpha_n|) a_3 \frac{1}{\sqrt{n^p}} + |\alpha_n| a_3 \frac{1}{\sqrt{n^p}} \\ &\leq (1 + 2|\alpha_n|) \frac{a_3}{\sqrt{n^p}}. \end{aligned}$$

Let  $a_4 = \sup_{n \geq \bar{k}} (1 + 2|\alpha_n|)a_3 > 0$ . Then,

$$\|y_n - \bar{x}\| \leq a_4 \frac{1}{\sqrt{n^p}}, \text{ for all } n \geq \bar{k}. \quad (31)$$

■

In case the Lojasiewicz exponent of the regularization function  $H$  is  $\theta \in (\frac{1}{2}, 1)$  we have the following result concerning the convergence rates of the sequences generated by (1).

**Theorem 14** *In the settings of problem (P) consider the sequences  $(x_n)_{n \in \mathbb{N}}$ ,  $(y_n)_{n \in \mathbb{N}}$  generated by Algorithm (1). Assume that  $g$  is bounded from below and that  $(x_n)_{n \in \mathbb{N}}$  is bounded, let  $\bar{x} \in \text{crit}(g)$  be such that  $\lim_{n \rightarrow +\infty} x_n = \bar{x}$  and suppose that*

$$H : \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}, H(x, y) = g(x) + M\|x - y\|^2$$

*fulfills the Lojasiewicz property at  $(\bar{x}, \bar{x}) \in \text{crit } H$  with Lojasiewicz exponent  $\theta \in (\frac{1}{2}, 1)$ . Then, there exist  $b_1, b_2, b_3, b_4 > 0$  such that the following statements hold true:*

$$(b_1) \quad g(y_n) - g(\bar{x}) \leq b_1 \frac{1}{n^{\frac{1}{2\theta-1}}}, \text{ for all } n \geq \bar{N}_1 + 2;$$

$$(b_2) \quad g(x_n) - g(\bar{x}) \leq b_2 \frac{1}{n^{\frac{1}{2\theta-1}}}, \text{ for all } n \geq \bar{N}_1 + 2;$$

$$(b_3) \quad \|x_n - \bar{x}\| \leq b_3 \frac{1}{n^{\frac{1-\theta}{2\theta-1}}}, \text{ for all } n \geq \bar{N}_1 + 2;$$

$$(b_4) \quad \|y_n - \bar{x}\| \leq b_4 \frac{1}{n^{\frac{1-\theta}{2\theta-1}}}, \text{ for all } n > \bar{N}_1 + 2,$$

where  $\bar{N}_1 \in \mathbb{N}$  was defined in the proof of Theorem 13.

**Proof.** To avoid triviality also here we assume that  $H_n > \bar{H}$  for all  $n \geq N$ .

We return to (22) which, after using  $r_{n+1} \leq r_n$ , we rewrite as

$$(r_n - r_{n+1})r_{n+1}^{-2\theta} + \mu_n(r_{n+1} - r_{n+2})r_{n+1}^{-2\theta} \geq \lambda_n, \quad (32)$$

for all  $n \geq \bar{N}_1$ .

Now, considering the real-valued function  $\phi(t) = \frac{K}{2\theta-1}t^{1-2\theta}$  whose derivative is  $\phi'(t) = -Kt^{-2\theta}$ , one has that

$$\phi(r_{n+1}) - \phi(r_n) = \int_{r_n}^{r_{n+1}} \phi'(t)dt \geq K(r_n - r_{n+1})r_n^{-2\theta}$$

and by similar arguments

$$\phi(r_{n+2}) - \phi(r_{n+1}) \geq K(r_{n+1} - r_{n+2})r_{n+1}^{-2\theta}.$$

Assume that for some  $n \geq \bar{N}_1$  one has

$$r_n^{-2\theta} \geq \frac{1}{2}r_{n+1}^{-2\theta}.$$

This leads to the following chain of inequalities :

$$\begin{aligned} \phi(r_{n+1}) - \phi(r_n) + \mu_n(\phi(r_{n+2}) - \phi(r_{n+1})) &\geq \frac{K}{2}(r_n - r_{n+1})r_{n+1}^{-2\theta} + K\mu_n(r_{n+1} - r_{n+2})r_{n+1}^{-2\theta} \\ &\geq \frac{K}{2}(r_n - r_{n+1})r_{n+1}^{-2\theta} + \frac{K}{2}\mu_n(r_{n+1} - r_{n+2})r_{n+1}^{-2\theta} \\ &\geq \frac{K}{2}\lambda_n. \end{aligned} \quad (33)$$



On the other hand, if  $2r_n^{-2\theta} < r_{n+1}^{-2\theta}$ , for some  $n \geq \bar{N}_1$ , then we obtain

$$2^{\frac{2\theta-1}{2\theta}} r_n^{1-2\theta} < r_{n+1}^{1-2\theta}.$$

As a consequence, it follows that

$$\phi(r_{n+1}) - \phi(r_n) = \frac{K}{2\theta-1} (r_{n+1}^{1-2\theta} - r_n^{1-2\theta}) \geq \frac{K}{2\theta-1} \left(2^{\frac{2\theta-1}{2\theta}} - 1\right) r_n^{1-2\theta} \geq \frac{K}{2\theta-1} \left(2^{\frac{2\theta-1}{2\theta}} - 1\right) r_{\bar{N}_1}^{1-2\theta} = C_1.$$

This implies that

$$\phi(r_{n+1}) - \phi(r_n) + \mu_n (\phi(r_{n+2}) - \phi(r_{n+1})) \geq C_1(1 + \mu_n).$$

Now, let us denote  $\inf_{n \geq \bar{N}_1} \frac{C_1(1 + \mu_n)}{\lambda_n}$  by  $C_2$ . Then

$$\phi(r_{n+1}) - \phi(r_n) + \mu_n (\phi(r_{n+2}) - \phi(r_{n+1})) \geq C_2 \lambda_n. \quad (34)$$

From the inequalities (33) and (34), it follows that there exist  $C > 0$ , such that

$$\phi(r_{n+1}) - \phi(r_n) + \mu_n (\phi(r_{n+2}) - \phi(r_{n+1})) \geq C \lambda_n, \quad \text{for every } n \geq \bar{N}_1.$$

By taking  $\bar{\mu} = \sup_{n \geq \bar{N}_1} \mu_n$  and summing over the indices we reach

$$\sum_{k=\bar{N}_1}^n ((\phi(r_{k+1}) - \phi(r_k)) + \bar{\mu}(\phi(r_{k+2}) - \phi(r_{k+1}))) \geq C \sum_{k=\bar{N}_1}^n \lambda_k.$$

Consequently

$$\phi(r_{n+1}) - \phi(r_{\bar{N}_1}) + \bar{\mu}(\phi(r_{n+2}) - \phi(r_{\bar{N}_1 k+1})) \geq C \sum_{k=\bar{N}_1}^n \lambda_k.$$

However, recall that both  $(r_n)_{n \geq \bar{N}_1}$  and the function  $\phi$  are decreasing, so we have

$$(1 + \bar{\mu})\phi(r_{n+2}) \geq C \sum_{k=\bar{N}_1}^n \lambda_k$$

and consequently

$$r_n^{1-2\theta} \geq \frac{C(2\theta-1)}{K(1+\bar{\mu})} \sum_{k=\bar{N}_1}^{n-2} \lambda_k,$$

for all  $n \geq \bar{N}_1 + 2$ . Therefore,

$$r_n \leq \left(\frac{C(2\theta-1)}{K(1+\bar{\mu})}\right)^{-\frac{1}{2\theta-1}} \left(\sum_{k=\bar{N}_1}^{n-2} \lambda_k\right)^{-\frac{1}{2\theta-1}} \quad \text{for all } n \geq \bar{N}_1 + 2.$$

The desired convergence rate, now follows from  $r_n = g(y_n) - g(\bar{x}) + \delta_n \|x_n - x_{n-1}\|^2$  and  $\sum_{k=\bar{N}_1}^{n-2} \lambda_k \geq \underline{\lambda}(n - \bar{N}_1 - 1)$ , where  $0 < \underline{\lambda} = \inf_{k \geq \bar{N}_1} \lambda_k$ . In fact, we have

$$\left(\sum_{k=\bar{N}_1}^{n-2} \lambda_k\right)^{-\frac{1}{2\theta-1}} \leq \underline{\lambda}^{-\frac{1}{2\theta-1}} (n - \bar{N}_1 - 1)^{-\frac{1}{2\theta-1}} \leq \underline{\lambda}^{-\frac{1}{2\theta-1}} M n^{-\frac{1}{2\theta-1}}$$

for some  $M > 0$  and for all  $n \geq \bar{N}_1 + 2$ . Let  $b_1 = \left(\frac{C(2\theta-1)}{K(1+\bar{\mu})}\right)^{-\frac{1}{2\theta-1}} \lambda^{-\frac{1}{2\theta-1}} M$ . Then,

$$g(y_n) - g(\bar{x}) \leq b_1 n^{\frac{-1}{2\theta-1}}, \text{ for all } n \geq \bar{N}_1 + 2.$$

The other claims now follow quite easily. Indeed, note that (27) holds for every  $n \geq \bar{N}_1$ , hence

$$g(x_n) - g(y_n) \leq \frac{1}{2D\beta_n(2 - \beta_n L_g)} r_{n+1} \leq \frac{1}{2D\beta_n(2 - \beta_n L_g)} b_1 (n+1)^{\frac{-1}{2\theta-1}}.$$

Therefore, one obtains

$$g(x_n) - g(\bar{x}) = (g(x_n) - g(y_n)) + (g(y_n) - g(\bar{x})) \leq \left(\frac{1}{2D\beta_n(2 - \beta_n L_g)} b_1 + b_1\right) n^{\frac{-1}{2\theta-1}} = b_2 n^{\frac{-1}{2\theta-1}},$$

or every  $n \geq \bar{N}_1 + 2$ .

For proving (b<sub>3</sub>), we use (29) again, and we have that for all  $n \geq \bar{N}_1 + 2$  it holds

$$\begin{aligned} \|x_n - \bar{x}\| &\leq M_1 r_n^{1-\theta} \leq M_1 (b_1 n^{\frac{-1}{2\theta-1}})^{1-\theta} \leq \\ &= b_3 n^{\frac{\theta-1}{2\theta-1}}. \end{aligned}$$

The final estimate is a straightforward consequence of the definition of  $y_n$  and the above estimates. Indeed, for all  $n > \bar{N}_1 + 2$  one has

$$\begin{aligned} \|y_n - \bar{x}\| &= \left\| x_n + \alpha_n(x_n - x_{n-1}) - \bar{x} \right\| \leq |1 + \alpha_n| \cdot \|x_n - \bar{x}\| + |\alpha_n| \cdot \|x_{n-1} - \bar{x}\| \\ &\leq (1 + |\alpha_n|) b_3 n^{\frac{\theta-1}{2\theta-1}} + |\alpha_n| b_3 (n-1)^{\frac{\theta-1}{2\theta-1}} \leq (1 + 2|\alpha_n|) b_3 (n-1)^{\frac{\theta-1}{2\theta-1}}. \end{aligned}$$

Let  $b_4 = \sup_{n \geq \bar{N}_1 + 3} (1 + 2|\alpha_n|) b_3 \left(\frac{n}{n-1}\right)^{\frac{1-\theta}{2\theta-1}} > 0$ . Then,

$$\|y_n - \bar{x}\| \leq b_4 n^{\frac{\theta-1}{2\theta-1}}, \text{ for all } n > \bar{N}_1 + 2. \quad (35)$$

■

**Remark 15** According to [16],  $H$  is KL with Łojasiewicz exponent  $\theta \in [\frac{1}{2}, 1)$ , whenever  $g$  is KL with Łojasiewicz exponent  $\theta \in [\frac{1}{2}, 1)$ . Therefore, in the hypotheses of Theorem 13 and Theorem 14 it is enough to assume that  $g$  is KL with Łojasiewicz exponent  $\theta \in [\frac{1}{2}, 1)$ .

## 4 Numerical experiments

The aim of this section is to support the analytic results of the previous sections by numerical experiments and to highlight some interesting features of the generic algorithm (1). To this end, let us consider

**Algorithm 1**

$$\begin{cases} x_{n+1} = y_n - s \nabla g(y_n), \\ y_n = x_n - \frac{0.1n}{n+3} (x_n - x_{n-1}), \end{cases} \quad (36)$$

which is a specific version of the generic algorithm (1) with a constant, small enough, step-size and an always negative coefficient, i.e.,

$$\beta_n = s \quad \text{and} \quad \alpha_n = -\frac{0.1n}{n+3} \quad \text{with} \quad \lim_{n \rightarrow +\infty} \alpha_n = -0.1 \in \left(\frac{-10 + \sqrt{68}}{8}, 0\right).$$

In order to give a better perspective on its advantages and disadvantages, we compare the proposed algorithm with:

- a) the standard Nesterov algorithm, in the case of convex objective functions;
- b) a Nesterov-like algorithm (see [15]), in the nonconvex case.

In this respect, it may be useful to recall that until recently little was known about the efficiency of Nesterov’s accelerated gradient method outside a convex setting. However, in [15], a Nesterov-like method, differing from the original only by a multiplicative coefficient, has been studied and convergence rates have been provided for the very general case of nonconvex coercive objective functions that have the KL property.

The algorithms, considered in comparison to Algorithm 1, are:

**A form of Nesterov’s accelerated gradient method**

$$\begin{cases} x_{n+1} = y_n - s\nabla g(y_n), \\ y_n = x_n + \frac{n}{n+3}(x_n - x_{n-1}), \end{cases} \quad (37)$$

**A Nesterov-like algorithm for nonconvex optimization**

$$\begin{cases} x_{n+1} = y_n - s\nabla g(y_n), \\ y_n = x_n + \frac{\gamma n}{n+3}(x_n - x_{n-1}), \quad \text{with } \gamma \in (0, 1). \end{cases} \quad (38)$$

We start our experiments by testing Algorithm 1 in the simplest case of a one-dimensional, quadratic function  $g(x) = x^2$ . Here, the comparison between Algorithm 1 and Nesterov’s accelerated gradient (see [18]), with both methods sharing the same step-size  $s = 0.1$  shows that, although initially slower, at some point Algorithm 1 surpasses Nesterov’s method, as can be observed in Fig. 1 a).

The same behaviour can be observed also in a higher-dimensional setting, e.g.,  $g(x_1, x_2) = 0.02x_1^2 + 0.005x_2^2$ , with the decay of both the energy error and the error in terms of iterates presented in Fig. 1 c) and d), respectively. The difference of order between the decay in terms of the objective function and that of the norm of the iterates proved in Theorem 13 is confirmed by these two pictures. On the other hand, Fig. 1 b) shows a comparison between the trajectories generated by the two algorithms, in the  $x_1x_2$ -plane.

A second set of numerical experiments is related to the minimization of the nonconvex, coercive function  $g : \mathbb{R} \rightarrow \mathbb{R}$

$$g(x) = \ln(1 + (x^2 - 1)^2)$$

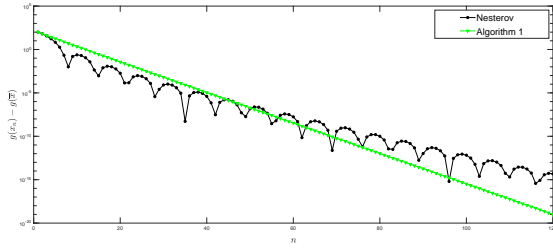
depicted in Fig. 2 a). When the starting point lies within the convexity domain of  $g$  (e.g.,  $x_0 = y_0 = 3$ , in Fig. 2 b)), Algorithm 1 converges faster than a Nesterov-like algorithm, after a certain number of iterates. However, outside the convexity domain of  $g$ , e.g., for  $x_0 = y_0 = 0.01$ , the Nesterov-like algorithm clearly outperforms Algorithm 1 (see Fig. 2 c)).

Nevertheless, the very general structure of the generic algorithm (1) allows for much flexibility, as only the limit of the sequence  $(\alpha_n)$  is prescribed. So, one can profit by switching between coefficients of different nature as is the case with the following modified version of Algorithm 1

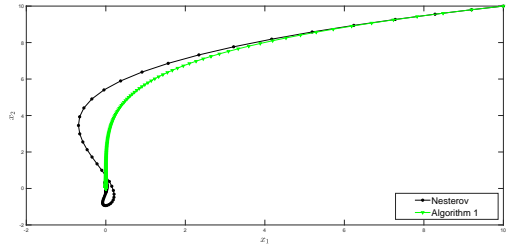
**Algorithm 2**

$$\begin{cases} x_{n+1} = y_n - s\nabla g(y_n), \\ y_n = x_n + \left( (1 - \lambda_n) \left( -0.1 \frac{n}{n+3} \right) + \lambda_n \frac{n}{n+3} \right) (x_n - x_{n-1}), \\ \lambda_n = \frac{1}{1 + \exp(1 + (5n - 20))}, \end{cases} \quad (39)$$

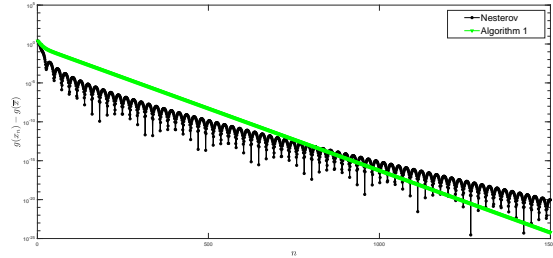
which first copies the Nesterov-like algorithm and then, after a number of iterates, switches to Algorithm 1, thus outperforming the Nesterov-like algorithm (see Fig. 2 d)).



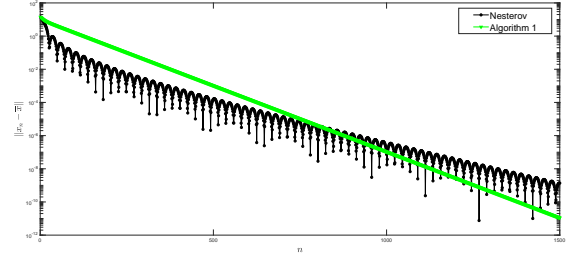
(a)  $g(x) = x^2, x_0 = 3, s = 2$



(b)  $g = 0.02x_1^2 + 0.005x_2^2, x_0 = (3, 1), s = 2$

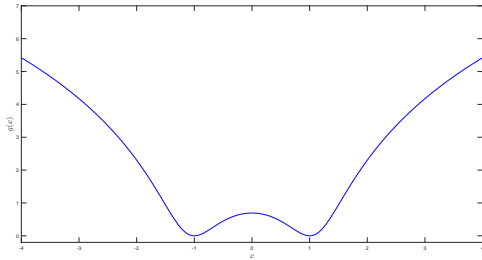


(c)  $g = 0.02x_1^2 + 0.005x_2^2, x_0 = (3, 1), s = 2$

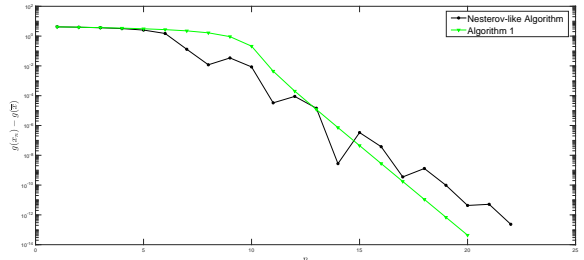


(d)  $g = 0.02x_1^2 + 0.005x_2^2, x_0 = (3, 1), s = 2$

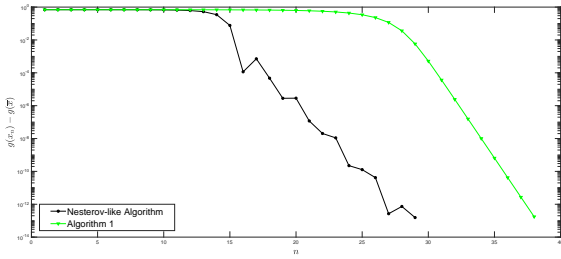
Figure 1: Minimizing quadratic functions in dimension one (in a) and two (in b-d)).



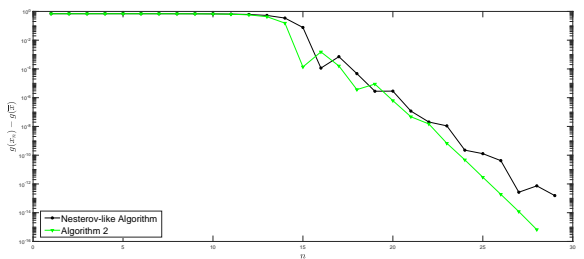
(a)  $g : \mathbb{R} \rightarrow \mathbb{R}, g(x) = \ln(1 + (x^2 - 1)^2)$



(b)  $x_0 = 3$  with  $s = 0.1$



(c)  $x_0 = 0.01$  with  $s = 0.1$



(d)  $x_0 = 0.01$  with  $s = 0.1$

Figure 2: Minimizing the nonconvex function  $g(x) = \ln(1 + (x^2 - 1)^2)$  by using different algorithms and different starting points.

## References

- [1] H. Attouch, J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Mathematical Programming 116(1-2) Series B, 5-16, 2009
- [2] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality*, Mathematics of Operations Research 35(2), 438-457, 2010
- [3] H. Attouch, J. Bolte, B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Mathematical Programming 137(1-2) Series A, 91-129, 2013
- [4] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Math. Program. 168(1-2) Ser. B, 123-175, 2018
- [5] P. Bégout, J. Bolte, M.A. Jendoubi, *On damped second-order gradient systems*, Journal of Differential Equations (259), 3115-3143, 2015
- [6] J. Bolte, S. Sabach, M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming Series A (146)(1-2), 459-494, 2014
- [7] J. Bolte, A. Daniilidis, A. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization 17(4), 1205-1223, 2006
- [8] J. Bolte, A. Daniilidis, A. Lewis, M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization 18(2), 556-572, 2007
- [9] J. Bolte, A. Daniilidis, O. Ley, L. Mazet, *Characterizations of Lojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society 362(6), 3319-3363, 2010
- [10] R.I. Boż, E.R. Csetnek, S.C. László, *Approaching nonsmooth nonconvex minimization through second-order proximal-gradient dynamical systems*, Journal of Evolution Equations 18(3), 1291-1318, 2018
- [11] R.I. Boż, E.R. Csetnek, S.C. László, *An inertial forward-backward algorithm for minimizing the sum of two non-convex functions*, Euro Journal on Computational Optimization 4(1), 3-25, 2016
- [12] R.I. Boż, E.R. Csetnek, S.C. László, *A second order dynamical approach with variable damping to nonconvex smooth minimization*, Applicable Analysis (2018), <https://doi.org/10.1080/00036811.2018.1495330>
- [13] P. Frankel, G. Garrigos, J. Peypouquet, *Splitting Methods with Variable Metric for KurdykaLojasiewicz Functions and General Convergence Rates*, Journal of Optimization Theory and Applications, 165(3), 874900, 2015
- [14] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Annales de l'institut Fourier (Grenoble) 48(3), 769-783, 1998
- [15] S.C. László, *Convergence rates for an inertial algorithm of gradient type associated to a smooth nonconvex minimization*, <https://arxiv.org/abs/1807.00387>

- [16] G. Li, T. K. Pong, *Calculus of the Exponent of Kurdyka-Lojasiewicz Inequality and Its Applications to Linear Convergence of First-Order Methods*, Foundations of Computational Mathematics, 1-34, 2018
- [17] S. Lojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique Paris, 87-89, 1963
- [18] Y.E. Nesterov, *A method for solving the convex programming problem with convergence rate  $O(1/k^2)$* , (Russian) Dokl. Akad. Nauk SSSR 269(3), 543-547, 1983
- [19] Y. Nesterov, *Introductory lectures on convex optimization: a basic course*. Kluwer Academic Publishers, Dordrecht, 2004
- [20] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, U.S.S.R. Comput. Math. Math. Phys., 4(5):1-17, 1964
- [21] R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis*, Fundamental Principles of Mathematical Sciences 317, Springer-Verlag, Berlin, 1998
- [22] W. Su, S. Boyd, E.J. Candes, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*, Journal of Machine Learning Research, 17, 1-43, 2016