Convergence rates for an inertial algorithm of gradient type associated to a smooth nonconvex minimization

Szilárd Csaba László [*]

November 22, 2018

Abstract. We investigate an inertial algorithm of gradient type in connection with the minimization of a nonconvex differentiable function. The algorithm is formulated in the spirit of Nesterov's accelerated convex gradient method. 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.

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

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

1 Introduction

Let $g : \mathbb{R}^m \longrightarrow \mathbb{R}$ be a (not necessarily convex) 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 $x, y \in \mathbb{R}^m$. We deal with the optimization problem

$$(P) \quad \inf_{x \in \mathbb{R}^m} g(x). \tag{1}$$

We associate to (1) the following inertial algorithm of gradient type. Consider the starting points $x_0 = y_0 \in \mathbb{R}^m$, and for all $n \in \mathbb{N}$

$$\begin{cases} x_{n+1} = y_n - s\nabla g(y_n), \\[2mm] y_n = x_n + \dfrac{\beta n}{n + \alpha}(x_n - x_{n-1}), \end{cases} \tag{2}$$

where $\alpha > 0, \beta \in (0, 1)$ and $0 < s < \frac{2(1-\beta)}{L_g}$.

Note that (2) is a nonconvex descendant of the methods of Polyak [23] and Nesterov [21]. Indeed, in [23], Polyak introduced a modified gradient method for minimizing a smooth convex function g. His two-step iterative method, the so called heavy ball method, takes the following form:

$$\begin{cases} x_{n+1} = y_n - \lambda_n \nabla g(x_n), \\[2mm] y_n = x_n + \alpha_n(x_n - x_{n-1}), \end{cases} \tag{3}$$

1

where $\alpha_n \in [0, 1)$ and $\lambda_n > 0$ is a step-size parameter.

In a seminal paper [21], Nesterov proposed a modification of the heavy ball method in order to obtain optimal convergence rates for smooth convex functions. More precisely, Nesterov used $\alpha_n = \frac{t_n - 1}{t_{n+1}}$ where t_n satisfies the recursion $t_{n+1} = \frac{\sqrt{4t_n^2 + 1} + 1}{2}$, $t_1 = 1$ and put y_n also for evaluating the gradient. Additionally, λ_n is chosen in such way that $\lambda_n \leq \frac{1}{L_g}$. His scheme in its simplest form is given by:

$$\begin{cases} x_{n+1} = y_n - s\nabla g(y_n), \\ \\ y_n = x_n + \dfrac{t_n - 1}{t_{n+1}}(x_n - x_{n-1}), \end{cases} \tag{4}$$

where $s \leq \frac{1}{L_g}$.

This scheme leads to the convergence rate $g(x_n) - g(\overline{x}) = \mathcal{O}\left(1/n^2\right)$, where $\overline{x}$ is a minimizer of the convex function g, and this is optimal among all methods having only information about the gradient of g and consecutive iterates, [22].

By taking $t_n = \frac{n + a - 1}{a}$, $a \geq 2$ in (4) we obtain an algorithm that is asymptotically equivalent to the original Nesterov method and leads to the same rate of convergence $\mathcal{O}\left(1/n^2\right)$, (see [15, 25]). This case has been considered by Chambolle and Dossal [15], in order to prove the convergence of the iterates of the modified FISTA algorithm (see [7]). We emphasize that Algorithm (2) has a similar form as the algorithm studied by Chambolle and Dossal (see [15] and also [18]), but we allow the function g to be nonconvex. Unfortunately, our analysis do not cover the case $\beta = 1$.

Su, Boyd and Candès (see [25]), showed that in case $t_n = \frac{n+1}{2}$ the algorithm (4) has the exact limit the second order differential equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla g(x(t)) = 0. \tag{5}$$

with $\alpha = 3$.

Recently, Attouch and his co-authors (see [4, 6]), proved that, if $\alpha > 3$ in (5), then the generated trajectory $x(t)$ converges to a minimizer of g as $t \longrightarrow +\infty$, while the convergence rate of the objective function along the trajectory is $o(1/t^2)$. Further, in [5], some results concerning the convergence rate of the objective function g along the trajectory generated by (5), in the subcritical case $\alpha \leq 3$, have been obtained. However, the convergence of the generated trajectories by (5) in case g is nonconvex is still an open question. Some important steps in this direction have been made in [14] (see also [12]), where convergence of the trajectories of a system, that can be viewed as a perturbation of (5), have been obtained in a nonconvex setting. More precisely in [14] is considered the system

$$\ddot{x}(t) + \left(\gamma + \frac{\alpha}{t}\right)\dot{x}(t) + \nabla g(x(t)) = 0, \tag{6}$$

and it is shown that the generated trajectory converges to a critical point of g, if a regularization of g satisfies the Kurdyka-Łojasiewicz property.

In what follows we show that by choosing appropriate values of β, the numerical scheme (2) has as the exact limit the continuous second order dynamical systems (5) studied in [4–6, 25], and also the continuous dynamical system (6) studied in [14]. We take to this end in (2) small step sizes and follow the same approach as Su, Boyd and Candès in [25], (see also [14]). For this purpose we rewrite (2) in the form

$$\frac{x_{n+1} - x_n}{\sqrt{s}} = \frac{\beta n}{n + \alpha} \cdot \frac{x_n - x_{n-1}}{\sqrt{s}} - \sqrt{s}\nabla g(y_n) \ \forall n \geq 1 \tag{7}$$

and introduce the *Ansatz* $x_n \approx x(n\sqrt{s})$ for some twice continuously differentiable function $x : [0, +\infty) \to \mathbb{R}^n$. We let $n = \frac{t}{\sqrt{s}}$ and get $x(t) \approx x_n$, $x(t + \sqrt{s}) \approx x_{n+1}$, $x(t - \sqrt{s}) \approx x_{n-1}$. Then, as the step size s goes

to zero, from the Taylor expansion of x we obtain

$$\frac{x_{n+1} - x_n}{\sqrt{s}} = \dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s})$$

and

$$\frac{x_n - x_{n-1}}{\sqrt{s}} = \dot{x}(t) - \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s}).$$

Further, since

$$\sqrt{s}\|\nabla g(y_n) - \nabla g(x_n)\| \leq \sqrt{s}L_g\|y_n - x_n\| = \sqrt{s}L_g\left|\frac{\beta n}{n+\alpha}\right|\|x_n - x_{n-1}\| = o(\sqrt{s}),$$

it follows $\sqrt{s}\nabla g(y_n) = \sqrt{s}\nabla g(x_n) + o(\sqrt{s})$. Consequently, (7) can be written as

$$\dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s}) =$$

$$\frac{\beta t}{t + \alpha\sqrt{s}}\left(\dot{x}(t) - \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s})\right) - \sqrt{s}\nabla g(x(t)) + o(\sqrt{s})$$

or, equivalently

$$(t + \alpha\sqrt{s})\left(\dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s})\right) =$$

$$\beta t\left(\dot{x}(t) - \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s})\right) - \sqrt{s}(t + \alpha\sqrt{s})\nabla g(x(t)) + o(\sqrt{s}).$$

Hence,

$$\frac{1}{2}\left(\alpha\sqrt{s} + (1+\beta)t\right)\ddot{x}(t)\sqrt{s} + \left((1-\beta)t + \alpha\sqrt{s}\right)\dot{x}(t) + \sqrt{s}(t + \alpha\sqrt{s})\nabla g(x(t)) = o(\sqrt{s}). \qquad (8)$$

Now, if we take $\beta = 1 - \gamma s < 1$ in (8) for some $\frac{1}{s} > \gamma > 0$, we obtain

$$\frac{1}{2}\left(\alpha\sqrt{s} + (2-\gamma s)t\right)\ddot{x}(t)\sqrt{s} + \left(\gamma st + \alpha\sqrt{s}\right)\dot{x}(t) + \sqrt{s}(t + \alpha\sqrt{s})\nabla g(x(t)) = o(\sqrt{s}).$$

After dividing by $\sqrt{s}$ and letting $s \to 0$, we obtain

$$t\ddot{x}(t) + \alpha\dot{x}(t) + t\nabla g(x(t)) = 0,$$

which, after division by t, gives (5), that is

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

Similarly, by taking $\beta = 1 - \gamma\sqrt{s} < 1$ in (8), for some $\frac{1}{\sqrt{s}} > \gamma > 0$, we obtain

$$\frac{1}{2}\left(\alpha\sqrt{s} + (2-\gamma\sqrt{s})t\right)\ddot{x}(t)\sqrt{s} + \left(\gamma\sqrt{s}t + \alpha\sqrt{s}\right)\dot{x}(t) + \sqrt{s}(t + \alpha\sqrt{s})\nabla g(x(t)) = o(\sqrt{s}).$$

After dividing by $\sqrt{s}$ and letting $s \to 0$, we get

$$t\ddot{x}(t) + (\gamma t + \alpha)\dot{x}(t) + t\nabla g(x(t)) = 0,$$

which, after division by t, gives (6), that is

$$\ddot{x}(t) + \left(\gamma + \frac{\alpha}{t}\right)\dot{x}(t) + \nabla g(x(t)) = 0.$$

3

Consequently, our numerical scheme (2) can be seen as the discrete counterpart of the continuous dynamical systems (5) and (6), in a full nonconvex setting.

The techniques for proving the convergence of (2) use the same main ingredients as other algorithms for nonconvex optimization problems involving KL functions. More precisely, in the next section, we show a sufficient decrease property for the iterates, which also ensures that the iterates gap belongs to l^2, further we show that the set of cluster points of the iterates is included in the set of critical points of the objective function, and, finally, we use the KL property of an appropriate regularization of the objective function in order to obtain that the iterates gap belongs to l^1, which implies the convergence of the iterates, see also [3, 8, 13]. Moreover, 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 (2), as well as for the function values $g(x_n)$, $g(y_n)$ in the terms of the Łojasiewicz exponent of g and a regularization of g, respectively (for some general results see [16]).

2   The convergence of the generated sequences

In this section we investigate the convergence of the proposed algorithm. We show that the sequences generated by the numerical scheme (2) converge to a critical point of the objective function g, provided the regularization of g, $H(x,y) = g(x) + \frac{1}{2}\|y - x\|^2$, is a KL function. The main tool in our forthcoming analysis is the so called descent lemma, see [22].

Lemma 1 *Let $g : \mathbb{R}^m \longrightarrow \mathbb{R}$ be Fréchet differentiable with L_g Lipschitz continuous gradient. Then*

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

Now we are able to obtain a decrease property for the iterates generated by (2).

Theorem 2 *In the settings of problem (1), for some starting points $x_0 = y_0 \in \mathbb{R}^m$ let $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ be the sequences generated by the numerical scheme (2). Consider the sequences*

$$A_{n-1} = \frac{2 - sL_g}{2s}\left(\frac{(1 + \beta)n + \alpha}{n + \alpha}\right)^2 - \frac{\beta n((1 + \beta)n + \alpha)}{s(n + \alpha)^2},$$

$$C_{n-1} = \frac{2 - sL_g}{2s}\frac{\beta n - \beta}{n + \alpha - 1}\frac{(1 + \beta)n + \alpha}{n + \alpha} - \frac{1}{2s}\frac{\beta n - \beta}{n + \alpha - 1}\frac{\beta n}{n + \alpha}$$

and

$$\delta_n = A_{n-1} - C_{n-1}$$

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

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

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

Assume that g is bounded from below. Then, the following statements hold.

(ii) The sequence $\big(g(y_n) + \delta_n\|x_n - x_{n-1}\|^2\big)_{n\in\mathbb{N}}$ is convergent;

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

4

Proof. From (2) we have $\nabla g(y_n) = \frac{1}{s}(y_n - x_{n+1})$, hence

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

Now, from Lemma 1 we obtain

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

consequently we have

$$g(y_{n+1}) - \frac{L_g}{2}\|y_{n+1} - y_n\|^2 \leq g(y_n) + \frac{1}{s}\langle y_n - x_{n+1}, y_{n+1} - y_n \rangle. \qquad (9)$$

Further,

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

and

$$y_{n+1} - x_{n+1} = \frac{\beta(n+1)}{n+\alpha+1}(x_{n+1} - x_n),$$

hence

$$g(y_{n+1}) + \left(\frac{1}{s} - \frac{L_g}{2}\right)\|y_{n+1} - y_n\|^2 \leq g(y_n) + \frac{\frac{\beta(n+1)}{n+\alpha+1}}{s}\langle x_{n+1} - x_n, y_{n+1} - y_n \rangle. \qquad (10)$$

Since

$$y_{n+1} - y_n = \frac{(1+\beta)n + \alpha + \beta + 1}{n+\alpha+1}(x_{n+1} - x_n) - \frac{\beta n}{n+\alpha}(x_n - x_{n-1}),$$

we have,

$$\|y_{n+1} - y_n\|^2 = \left\|\frac{(1+\beta)n + \alpha + \beta + 1}{n+\alpha+1}(x_{n+1} - x_n) - \frac{\beta n}{n+\alpha}(x_n - x_{n-1})\right\|^2 =$$

$$\left(\frac{(1+\beta)n + \alpha + \beta + 1}{n+\alpha+1}\right)^2 \|x_{n+1} - x_n\|^2 + \left(\frac{\beta n}{n+\alpha}\right)^2 \|x_n - x_{n-1}\|^2 -$$

$$2\frac{(1+\beta)n + \alpha + \beta + 1}{n+\alpha+1}\frac{\beta n}{n+\alpha}\langle x_{n+1} - x_n, x_n - x_{n-1} \rangle,$$

and

$$\langle x_{n+1} - x_n, y_{n+1} - y_n \rangle = \left\langle x_{n+1} - x_n, \frac{(1+\beta)n + \alpha + \beta + 1}{n+\alpha+1}(x_{n+1} - x_n) - \frac{\beta n}{n+\alpha}(x_n - x_{n-1}) \right\rangle =$$

$$\frac{(1+\beta)n + \alpha + \beta + 1}{n+\alpha+1}\|x_{n+1} - x_n\|^2 - \frac{\beta n}{n+\alpha}\langle x_{n+1} - x_n, x_n - x_{n-1} \rangle.$$

Replacing the above equalities in (10), we obtain

$$g(y_{n+1}) + \frac{(2 - sL_g)\left(\frac{(1+\beta)n+\alpha+\beta+1}{n+\alpha+1}\right)^2 - 2\frac{\beta(n+1)((1+\beta)n+\alpha+\beta+1)}{(n+\alpha+1)^2}}{2s}\|x_{n+1} - x_n\|^2 \leq$$

$$g(y_n) - \frac{(2 - sL_g)\left(\frac{\beta n}{n+\alpha}\right)^2}{2s}\|x_n - x_{n-1}\|^2 +$$

$$\frac{(2 - sL_g)\frac{\beta n}{n+\alpha}\frac{(1+\beta)n+\alpha+\beta+1}{n+\alpha+1} - \frac{\beta n}{n+\alpha}\frac{\beta(n+1)}{n+\alpha+1}}{s}\langle x_{n+1} - x_n, x_n - x_{n-1} \rangle.$$

For simplicity let

$$B_n = \frac{(2 - sL_g)\left(\frac{\beta n}{n + \alpha}\right)^2}{2s}$$

for all $n \in \mathbb{N}$.

Hence we have

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

By using the equality

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

we obtain

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

Note that $A_n - C_n = \delta_{n+1}$ and let us denote $\Delta_n = B_n + A_{n-1} - C_{n-1} - C_n$. Consequently the following inequality holds.

$$C_n\|x_{n+1} + x_{n-1} - 2x_n\|^2 + \Delta_n\|x_n - x_{n-1}\|^2 \leq (g(y_n) + \delta_n\|x_n - x_{n-1}\|^2) - (g(y_{n+1}) + \delta_{n+1}\|x_{n+1} - x_n\|^2). \tag{12}$$

Since $0 < \beta < 1$ and $s < \frac{2(1-\beta)}{L_g}$, we have

$$\lim_{n \to +\infty} A_n = \frac{(2 - sL_g)(\beta + 1)^2 - 2\beta - 2\beta^2}{2s} > 0,$$

$$\lim_{n \to +\infty} B_n = \frac{(2 - sL_g)\beta^2}{2s} > 0,$$

$$\lim_{n \to +\infty} C_n = \frac{(2 - sL_g)(\beta^2 + \beta) - \beta^2}{2s} > 0,$$

$$\lim_{n \to +\infty} \Delta_n = \frac{2 - sL_g - 2\beta}{2s} > 0,$$

and

$$\lim_{n \to +\infty} \delta_n = \frac{2 - \beta^2 - sL_g(\beta + 1)}{2s} > 0.$$

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

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

which, in the view of (12), shows (i), that is, the sequence $g(y_n) + \delta_n\|x_n - x_{n-1}\|^2$ is decreasing for $n \geq N$.

Assume now that g is bounded from below. By using (12) again, we obtain

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

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

$$0 \leq D\|x_n - x_{n-1}\|^2 \leq (g(y_n) + \delta_n\|x_n - x_{n-1}\|^2) - (g(y_{n+1}) + \delta_{n+1}\|x_{n+1} - x_n\|^2), \tag{13}$$

for all $n \geq N$. Let $r > N$. By summing up the latter relation we have

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

which leads to

$$g(y_{r+1}) + D \sum_{n=N}^{r} \|x_n - x_{n-1}\|^2 \leq g(y_N) + \delta_N \|x_N - x_{N-1}\|^2. \tag{14}$$

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

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

which proves (iii).

The latter relation also shows that

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

hence

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

But then, from the fact that g is bounded from below we obtain that the sequence $g(y_n) + \delta_n \|x_n - x_{n-1}\|^2$ is bounded from below. On the other hand, from (i) we have that the sequence $g(y_n) + \delta_n \|x_n - x_{n-1}\|^2$ is decreasing for $n \geq N$, hence there exists

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

∎

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

$$\lim_{n \longrightarrow +\infty} (x_n - x_{n-1}) = 0. \tag{15}$$

Let us denote by $\omega((x_n)_{n \in \mathbb{N}})$ the set of cluster points of the sequence $(x_n)_{n \in \mathbb{N}}$, and denote by $\mathrm{crit}(g) = \{x \in \mathbb{R}^m : \nabla g(x) = 0\}$ the set of critical points of g.

In the following result we use the distance function to a set, defined for $A \subseteq \mathbb{R}^n$ as $\mathrm{dist}(x, A) = \inf_{y \in A} \|x - y\|$ for all $x \in \mathbb{R}^n$.

Lemma 4 *In the settings of problem (1), for some starting points $x_0 = y_0 \in \mathbb{R}^m$ consider the sequences $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$ generated by Algorithm (2). Assume that g is bounded from below and consider the function*

$$H : \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}, \, H(x, y) = g(x) + \frac{1}{2}\|y - x\|^2.$$

Consider further the sequence

$$u_n = \sqrt{2\delta_n}(x_n - x_{n-1}) + y_n, \, \text{for all } n \in \mathbb{N},$$

where δ_n was defined in Theorem 2. Then, the following statements hold true.

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

(ii) There exists and is finite the limit $\lim_{n \longrightarrow +\infty} H(y_n, u_n)$;

(iii) $\omega((y_n, u_n)_{n \in \mathbb{N}}) \subseteq \mathrm{crit}\, H = \{(x, x) \in \mathbb{R}^m \times \mathbb{R}^m : x \in \mathrm{crit}\, g\}$;

(iv) $\|\nabla H(y_n, u_n)\| \leq \frac{1}{s}\|x_{n+1} - x_n\| + \left(\frac{\beta n}{s(n+\alpha)} + 2\sqrt{2\delta_n}\right)\|x_n - x_{n-1}\|$ for all $n \in \mathbb{N}$;

7

(v) $\|\nabla H(y_n, u_n)\|^2 \le \frac{2}{s^2}\|x_{n+1} - x_n\|^2 + 2\left(\left(\frac{\beta n}{s(n+\alpha)} - \sqrt{2\delta_n}\right)^2 + \delta_n\right)\|x_n - x_{n-1}\|^2$ for all $n \in \mathbb{N}$;

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

Assume that $(x_n)_{n\in\mathbb{N}}$ is bounded. Then,

(vii) $\omega((y_n, u_n)_{n\in\mathbb{N}})$ is nonempty and compact;

(viii) $\lim_{n\longrightarrow+\infty} \operatorname{dist}((y_n, u_n), \omega((y_n, u_n)_{n\in\mathbb{N}})) = 0$.

Proof. (i) Let $\overline{x} \in \omega((x_n)_{n\in\mathbb{N}})$. Then, there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that

$$\lim_{k\to+\infty} x_{n_k} = \overline{x}.$$

Since by (15) $\lim_{n\longrightarrow+\infty}(x_n - x_{n-1}) = 0$ and the sequences $(\sqrt{2\delta_n})_{n\in\mathbb{N}}$, $\left(\frac{\beta n}{n+\alpha}\right)_{n\in\mathbb{N}}$ converge, we obtain that

$$\lim_{k\to+\infty} y_{n_k} = \lim_{k\to+\infty} u_{n_k} = \lim_{k\to+\infty} x_{n_k} = \overline{x},$$

which shows that

$$\omega((x_n)_{n\in\mathbb{N}}) \subseteq \omega((u_n)_{n\in\mathbb{N}}) \text{ and } \omega((x_n)_{n\in\mathbb{N}}) \subseteq \omega((y_n)_{n\in\mathbb{N}}).$$

Further from (2), the continuity of ∇g and (15), we obtain that

$$\nabla g(\overline{x}) = \lim_{k\longrightarrow+\infty} \nabla g(y_{n_k}) = \frac{1}{s}\lim_{k\longrightarrow+\infty}(y_{n_k} - x_{n_k+1}) =$$

$$\frac{1}{s}\lim_{k\longrightarrow+\infty}\left[(x_{n_k} - x_{n_k+1}) + \frac{\beta n_k}{n_k + \alpha}(x_{n_k} - x_{n_k-1})\right] = 0.$$

Hence $\omega((x_n)_{n\in\mathbb{N}}) \subseteq \operatorname{crit} g$. Conversely, if $\overline{u} \in \omega((u_n)_{n\in\mathbb{N}})$ then, from (15) results that $\overline{u} \in \omega((y_n)_{n\in\mathbb{N}})$ and $\overline{u} \in \omega((x_n)_{n\in\mathbb{N}})$. Hence,

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

(ii) is nothing else than (ii) in Theorem 2.

For (iii) observe that $\nabla H(x, y) = (\nabla g(x) + x - y, y - x)$, hence, $\nabla H(x, y) = 0$ leads to $x = y$ and $\nabla g(x) = 0$. Consequently

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

Further, consider $(\overline{y}, \overline{u}) \in \omega((y_n, u_n)_{n\in\mathbb{N}})$. Then, there exists $(y_{n_k}, u_{n_k})_{k\in\mathbb{N}} \subseteq (y_n, u_n)_{n\in\mathbb{N}}$ such that

$$(\overline{y}, \overline{u}) = \lim_{k\longrightarrow+\infty}(y_{n_k}, u_{n_k}) = \lim_{k\longrightarrow+\infty}(x_{n_k}, x_{n_k}) = (\overline{x}, \overline{x}).$$

Hence, $\overline{u} = \overline{y} = \overline{x} \in \omega((x_n)_{n\in\mathbb{N}}) \subseteq \operatorname{crit} g$ and $(\overline{x}, \overline{x}) \in \operatorname{crit} H$.

(iv) By using the 1-norm of $\mathbb{R}^m \times \mathbb{R}^m$ and (2), for every $n \in \mathbb{N}$ we have

$$\|\nabla H(y_n, u_n)\| \le \|\nabla H(y_n, u_n)\|_1 = \|(\nabla g(y_n) + y_n - u_n, u_n - y_n)\|_1 = \|\nabla g(y_n) + y_n - u_n\| + \|u_n - y_n\| \le$$

$$\|\nabla g(y_n)\| + 2\|\sqrt{2\delta_n}(x_n - x_{n-1})\| =$$

$$\frac{1}{s}\left\|\left(x_n + \frac{\beta n}{n + \alpha}(x_n - x_{n-1})\right) - x_{n+1}\right\| + 2\sqrt{2\delta_n}\|x_n - x_{n-1}\| \le$$

$$\frac{1}{s}\|x_{n+1} - x_n\| + \left(\frac{\beta n}{s(n + \alpha)} + 2\sqrt{2\delta_n}\right)\|x_n - x_{n-1}\|.$$

8

(v) We use the euclidian norm of $\mathbb{R}^m \times \mathbb{R}^m$, that is $\|(x,y)\| = \sqrt{\|x\|^2 + \|y\|^2}$ for all $(x,y) \in \mathbb{R}^m \times \mathbb{R}^m$. We have:

$$\|\nabla H(y_n, u_n)\|^2 = \|(\nabla g(y_n) + y_n - u_n, u_n - y_n)\|^2 = \|\nabla g(y_n) + y_n - u_n\|^2 + \|u_n - y_n\|^2 =$$

$$\left\| \frac{1}{s}(x_n - x_{n+1}) + \left(\frac{\beta n}{s(n+\alpha)} - \sqrt{2\delta_n} \right)(x_n - x_{n-1}) \right\|^2 + 2\delta_n \|x_n - x_{n-1}\|^2 \leq$$

$$\frac{2}{s^2}\|x_{n+1} - x_n\|^2 + 2\left(\left(\frac{\beta n}{s(n+\alpha)} - \sqrt{2\delta_n} \right)^2 + \delta_n \right) \|x_n - x_{n-1}\|^2$$

for all $n \in \mathbb{N}$.

(vi) follows directly from (ii).

Assume now that $(x_n)_{n \in \mathbb{N}}$ is bounded and let us prove (vii), (see also [14]). Obviously $(y_n, u_n)_{n \in \mathbb{N}}$ is bounded, hence according to Weierstrass Theorem $\omega((y_n, u_n)_{n \in \mathbb{N}})$, (and also $\omega((x_n)_{n \in \mathbb{N}})$), is nonempty. It remains to show that $\omega((y_n, u_n)_{n \in \mathbb{N}})$ is closed. From (i) and the proof of (iii) we have

$$\omega((y_n, u_n)_{n \in \mathbb{N}}) = \{(\overline{x}, \overline{x}) \in \mathbb{R}^m \times \mathbb{R}^m : \overline{x} \in \omega((x_n)_{n \in \mathbb{N}})\}. \tag{16}$$

Hence, it is enough to show that $\omega((x_n)_{n \in \mathbb{N}})$ is closed.

Let be $(\overline{x}_p)_{p \in \mathbb{N}} \subseteq \omega((x_n)_{n \in \mathbb{N}})$ and assume that $\lim_{p \longrightarrow +\infty} \overline{x}_p = x^*$. We show that $x^* \in \omega((x_n)_{n \in \mathbb{N}})$. Obviously, for every $p \in \mathbb{N}$ there exists a sequence of natural numbers $n_k^p \longrightarrow +\infty$, $k \longrightarrow +\infty$, such that

$$\lim_{k \longrightarrow +\infty} x_{n_k^p} = \overline{x}_p.$$

Let be $\epsilon > 0$. Since $\lim_{p \longrightarrow +\infty} \overline{x}_p = x^*$, there exists $P(\epsilon) \in \mathbb{N}$ such that for every $p \geq P(\epsilon)$ it holds

$$\|\overline{x}_p - x^*\| < \frac{\epsilon}{2}.$$

Let $p \in \mathbb{N}$ be fixed. Since $\lim_{k \longrightarrow +\infty} x_{n_k^p} = \overline{x}_p$, there exists $k(p, \epsilon) \in \mathbb{N}$ such that for every $k \geq k(p, \epsilon)$ it holds

$$\|x_{n_k^p} - \overline{x}_p\| < \frac{\epsilon}{2}.$$

Let be $k_p \geq k(p, \varepsilon)$ such that $n_{k_p}^p > p$. Obviously $n_{k_p}^p \longrightarrow \infty$ as $p \longrightarrow +\infty$ and for every $p \geq P(\epsilon)$

$$\|x_{n_{k_p}^p} - x^*\| < \epsilon.$$

Hence

$$\lim_{p \longrightarrow +\infty} x_{n_{k_p}^p} = x^*,$$

thus $x^* \in \omega((x_n)_{n \in \mathbb{N}})$.

(viii) By using (16) we have

$$\lim_{n \longrightarrow +\infty} \text{dist}((y_n, u_n), \omega((y_n, u_n)_{n \in \mathbb{N}})) = \lim_{n \longrightarrow +\infty} \inf_{\overline{x} \in \omega((x_n)_{n \in \mathbb{N}})} \|(y_n, u_n) - (\overline{x}, \overline{x})\|.$$

Since there exists the subsequences $(y_{n_k})_{k \in \mathbb{N}}$ and $(u_{n_k})_{k \in \mathbb{N}}$ such that $\lim_{k \longrightarrow \infty} y_{n_k} = \lim_{k \longrightarrow \infty} u_{n_k} = \overline{x}_0 \in \omega((x_n)_{n \in \mathbb{N}})$ it is straightforward that

$$\lim_{n \longrightarrow +\infty} \text{dist}((y_n, u_n), \omega((y_n, u_n)_{n \in \mathbb{N}})) = 0.$$

∎

Remark 5 We emphasize that if g is coercive, that is $\lim_{\|x\| \to +\infty} g(x) = +\infty$, then g is bounded from below and $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$, the sequences generated by (2), are bounded.

Indeed, notice that g is bounded from below, being a continuous and coercive function (see [24]). Note that according to Theorem 2 the sequence $D \sum_{n=N}^r \|x_n - x_{n-1}\|^2$ is convergent hence is bounded. Consequently, from (14) it follows that y_r is contained for every $r > N$, (N is defined in the hypothesis of Theorem 2), in a lower level set of g, which is bounded. Since $(y_n)_{n \in \mathbb{N}}$ is bounded, taking into account (15), it follows that $(x_n)_{n \in \mathbb{N}}$ is also bounded.

In order to continue our analysis we need the concept of a KL function. For $\eta \in (0, +\infty]$, we denote by Θ_η the class of concave and continuous functions $\varphi : [0, \eta) \to [0, +\infty)$ such that $\varphi(0) = 0$, φ is continuously differentiable on $(0, \eta)$, continuous at 0 and $\varphi'(s) > 0$ for all $s \in (0, \eta)$.

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

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

the following inequality holds

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

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

The origins of this notion go back to the pioneering work of Łojasiewicz [19], where it is proved that for a real-analytic function $f : \mathbb{R}^n \to \mathbb{R}$ and a critical point $\overline{x} \in \mathbb{R}^n$ (that is $\nabla f(\overline{x}) = 0$), there exists $\theta \in [1/2, 1)$ such that the function $|f - f(\overline{x})|^\theta \|\nabla f\|^{-1}$ is bounded around $\overline{x}$. This corresponds to the situation when $\varphi(s) = C(1 - \theta)^{-1} s^{1-\theta}$. The result of Łojasiewicz allows the interpretation of the KL property as a re-parametrization of the function values in order to avoid flatness around the critical points. Kurdyka [17] extended this property to differentiable functions definable in an o-minimal structure. Further extensions to the nonsmooth setting can be found in [2, 9–11].

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 [1–3, 8–11] and the references therein for more details regarding all the classes mentioned above and illustrating examples.

An important role in our convergence analysis will be played by the following uniformized KL property given in [8, Lemma 6].

Lemma 6 *Let $\Omega \subseteq \mathbb{R}^n$ be a compact set and let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Assume that f is constant on Ω and f satisfies the KL property at each point of Ω. Then there exist $\varepsilon, \eta > 0$ and $\varphi \in \Theta_\eta$ such that for all $\overline{x} \in \Omega$ and for all x in the intersection*

$$\{x \in \mathbb{R}^n : \text{dist}(x, \Omega) < \varepsilon\} \cap \{x \in \mathbb{R}^n : f(\overline{x}) < f(x) < f(\overline{x}) + \eta\} \tag{17}$$

the following inequality holds

$$\varphi'(f(x) - f(\overline{x}))\|\nabla f(x)\| \geq 1. \tag{18}$$

The following convergence result is the first main result of the paper.

Theorem 7 *In the settings of problem (1), for some starting points $x_0 = y_0 \in \mathbb{R}^m$ consider the sequences $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$ generated by Algorithm (2). Assume that g is bounded from below and consider the function*

$$H : \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}, \, H(x, y) = g(x) + \frac{1}{2}\|y - x\|^2.$$

Assume that $(x_n)_{n \in \mathbb{N}}$ is bounded and H is a KL function. Then the following statements are true

(a) $\sum_{n\geq 1} \|x_n - x_{n-1}\| < +\infty$;

(b) there exists $x \in \mathrm{crit}(g)$ such that $\lim_{n\longrightarrow+\infty} x_n = x$.

Proof. Consider the sequence

$$u_n = \sqrt{2\delta_n}(x_n - x_{n-1}) + y_n, \text{ for all } n \in \mathbb{N},$$

that was defined in the hypotheses of Lemma 4. Furthermore, consider $(\overline{x}, \overline{x}) \in \omega((y_n, u_n)_{n\in\mathbb{N}})$.

Then, according to Lemma 4, the sequence $H(y_n, u_n)$ is decreasing for all $n \geq N$, where N was defined in Theorem 2, further

$$\overline{x} \in \mathrm{crit}\, g \text{ and } \lim_{n\longrightarrow+\infty} H(y_n, u_n) = H(\overline{x}, \overline{x}).$$

We divide the proof into two cases.

Case I. There exists $\overline{n} \geq N$, $\overline{n} \in \mathbb{N}$, such that $H(y_{\overline{n}}, u_{\overline{n}}) = H(\overline{x}, \overline{x})$. Then, since $H(y_n, u_n)$ is decreasing for all $n \geq N$ and $\lim_{n\longrightarrow+\infty} H(y_n, u_n) = H(\overline{x}, \overline{x})$ we obtain that

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

The latter relation combined with (13) leads to

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

for all $n \geq \overline{n}$.

Hence $(x_n)_{n\geq\overline{n}}$ is constant and the conclusion follows.

Case II. For every $n \geq N$ one has that $H(y_n, u_n) > H(\overline{x}, \overline{x})$. Let $\Omega = \omega((y_n, u_n)_{n\in\mathbb{N}})$. Then according to Lemma 4, Ω is nonempty and compact and H is constant on Ω. Since H is KL, according to Lemma 6 there exist $\varepsilon, \eta > 0$ and $\varphi \in \Theta_\eta$ such that for all (z, w) belonging to the intersection

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

one has

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

Since $\lim_{n\longrightarrow+\infty} \mathrm{dist}((y_n, u_n), \Omega) = 0$, there exists $n_1 \in \mathbb{N}$ such that

$$\mathrm{dist}((y_n, u_n), \Omega) < \epsilon, \forall n \geq n_1.$$

Since

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

and

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

there exists $n_2 \geq N$ such that

$$H(\overline{x}, \overline{x}) < H(y_n, u_n) < H(\overline{x}, \overline{x}) + \eta, \forall n \geq n_2.$$

Hence, for all $n \geq \overline{n} = \max(n_1, n_2)$ we have

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

Since φ is concave, for all $n \in \mathbb{N}$ we have

$$\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\overline{x}, \overline{x})) \geq$$

11

$$\varphi'(H(y_n, u_n) - H(\overline{x}, \overline{x})) \cdot (H(y_n, u_n) - H(y_{n+1}, u_{n+1})),$$

hence,

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

for all $n \geq \overline{n}$.

Now, from (13) and Lemma 4 (iv) we obtain

$$\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\overline{x}, \overline{x})) \geq \tag{19}$$

$$\frac{D\|x_n - x_{n-1}\|^2}{\frac{1}{s}\|x_{n+1} - x_n\| + \left(\frac{\beta n}{s(n+\alpha)} + 2\sqrt{2\delta_n}\right)\|x_n - x_{n-1}\|},$$

for all $n \geq \overline{n}$. Since $\lim_{n \longrightarrow +\infty} \delta_n = \frac{2 - \beta^2 - sL_g(\beta+1)}{2s} > 0$ and $\lim_{n \longrightarrow +\infty} \frac{\beta n}{s(n+\alpha)} = \frac{\beta}{s} \geq 0$ there exists $\overline{N} \in \mathbb{N}$, $\overline{N} \geq \overline{n}$ and $M > 0$ such that

$$\max\left(\frac{1}{s}, \frac{\beta n}{s(n+\alpha)} + 2\sqrt{2\delta_n}\right) \leq M,$$

for all $n \geq \overline{N}$. Hence, (19) becomes

$$\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\overline{x}, \overline{x})) \geq \tag{20}$$

$$\frac{D\|x_n - x_{n-1}\|^2}{M(\|x_{n+1} - x_n\| + \|x_n - x_{n-1}\|)},$$

for all $n \geq \overline{N}$.

Consequently,

$$\|x_n - x_{n-1}\| \leq$$

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

for all $n \geq \overline{N}$.

By using the arithmetical-geometrical mean inequality we have

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

$$\frac{\|x_{n+1} - x_n\| + \|x_n - x_{n-1}\|}{3} + \frac{3M}{4D}(\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\overline{x}, \overline{x})))$$

for all $n \geq \overline{N}$. Hence

$$\|x_n - x_{n-1}\| \leq$$

$$\frac{\|x_{n+1} - x_n\| + \|x_n - x_{n-1}\|}{3} + \frac{3M}{4D}(\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\overline{x}, \overline{x})))$$

for all $n \geq \overline{N}$ which leads to

$$2\|x_n - x_{n-1}\| - \|x_{n+1} - x_n\| \leq \frac{9M}{4D}(\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})) - \varphi(H(y_{n+1}, u_{n+1}) - H(\overline{x}, \overline{x}))) \tag{21}$$

for all $n \geq \overline{N}$. Let $P > \overline{N}$. By summing up (21) from $\overline{N}$ to P we obtain

$$\sum_{n=\overline{N}}^{P} \|x_n - x_{n-1}\| \leq$$

$$-\|x_{\overline{N}} - x_{\overline{N}-1}\| + \|x_{P+1} - x_P\| + \frac{9M}{4D}(\varphi(H(y_{\overline{N}}, u_{\overline{N}}) - H(\overline{x}, \overline{x})) - \varphi(H(y_{P+1}, u_{P+1}) - H(\overline{x}, \overline{x}))).$$

Now, by letting $P \longrightarrow +\infty$ and using the fact that $\varphi(0) = 0$ and (15) we obtain that

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

hence

$$\sum_{n \geq 1} \|x_n - x_{n-1}\| < +\infty$$

which is exactly (a).

Obviously the sequence $S_n = \sum_{k=1}^{n} \|x_k - x_{k-1}\|$ is Cauchy, hence, for all $\epsilon > 0$ there exists $N_\epsilon \in \mathbb{N}$ such that for all $n \geq N_\epsilon$ and for all $p \in \mathbb{N}$ one has

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

But

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

hence the sequence $(x_n)_{n \in \mathbb{N}}$ is Cauchy, consequently is convergent. Let

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

Now, according to Lemma 4 (i) one has

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

which proves (b). ∎

Remark 8 Since the class of semi-algebraic functions is closed under addition (see for example [8]) and $(x, y) \mapsto \frac{1}{2}\|x - y\|^2$ 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.

Remark 9 Note that, according to Remark 5, the conclusion of Theorem 7 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.

Remark 10 Note that under the assumptions of Theorem 7 we have $\lim_{n \longrightarrow +\infty} y_n = x$ and

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

3 Convergence rates

In this section we will assume that the regularized function H satisfies the Łojasiewicz property, which, as noted in the previous section, corresponds to a particular choice of the desingularizing function φ (see [1, 9, 19]).

Definition 2 *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a differentiable function. The function f is said to fulfill the Łojasiewicz property, if for every $\overline{x} \in \operatorname{crit} f$ there exist $K, \epsilon > 0$ and $\theta \in (0, 1)$ such that*

$$|f(x) - f(\overline{x})|^{\theta} \le K\|\nabla f(x)\| \text{ for every } x \text{ fulfilling } \|x - \overline{x}\| < \epsilon.$$

The number θ is called the Łojasiewicz exponent of f at the critical point $\overline{x}$. This corresponds to the case when the desingularizing function φ has the form $\varphi(t) = \frac{K}{1-\theta} t^{1-\theta}$.

In the following theorems we provide convergence rates for the sequence generated by (2), but also for the function values, in terms of the Łojasiewicz exponent of H (see, also, [1, 9]). 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 11 *In the settings of problem (1) consider the sequences $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$ generated by Algorithm (2). Assume that g is bounded from below and that $(x_n)_{n \in \mathbb{N}}$ is bounded, let $\overline{x} \in \operatorname{crit}(g)$ be such that $\lim_{n \longrightarrow +\infty} x_n = \overline{x}$ and suppose that*

$$H : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}, \ H(x, y) = g(x) + \frac{1}{2}\|x - y\|^2$$

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

(a_1) $g(y_n) - g(\overline{x}) \le a_1 \frac{1}{n^p}$ *for every* $n > \overline{k}$,

(a_2) $g(x_n) - g(\overline{x}) \le a_2 \frac{1}{n^p}$ *for every* $n > \overline{k}$,

(a_3) $\|x_n - \overline{x}\| \le a_3 \frac{1}{n^{\frac{p}{2}}}$ *for every* $n > \overline{k}$,

(a_4) $\|y_n - \overline{x}\| \le a_4 \frac{1}{n^{\frac{p}{2}}}$ *for all* $n > \overline{k}$.

Proof. As we have seen in the proof of Theorem 7, if there exists $\overline{n} \ge N$, $\overline{n} \in \mathbb{N}$, (where N was defined in Theorem 2), such that $H(y_{\overline{n}}, u_{\overline{n}}) = H(\overline{x}, \overline{x})$, then,

$$H(y_n, u_n) = H(\overline{x}, \overline{x}) \text{ for all } n \ge \overline{n}$$

and $(x_n)_{n \ge \overline{n}}$ is constant. Consequently $(y_n)_{n \ge \overline{n}}$ is also constant and the conclusion of the theorem is straightforward.

Hence, in what follows we assume that $H(y_n, u_n) > H(\overline{x}, \overline{x})$, for all $n \ge N$.

Let us fix $p > 0$ and let us prove (a_1).

For simplicity let us denote $r_n = H(y_n, u_n) - H(\overline{x}, \overline{x}) > 0$ for all $n \in \mathbb{N}$. From (12) we have

$$\Delta_n \|x_n - x_{n-1}\|^2 \le r_n - r_{n+1} \text{ for all } n \ge N.$$

From Lemma 4 (v) we have

$$\|\nabla H(y_n, u_n)\|^2 \le \frac{2}{s^2}\|x_{n+1} - x_n\|^2 + 2\left(\left(\frac{\beta n}{s(n + \alpha)} - \sqrt{2\delta_n}\right)^2 + \delta_n\right)\|x_n - x_{n-1}\|^2$$

14

for all $n \in \mathbb{N}$. Let $S_n = 2 \left(\left(\frac{\beta_n}{s(n+\alpha)} - \sqrt{2\delta_n} \right)^2 + \delta_n \right)$, for all $n \in \mathbb{N}$.

It follows that, for all $n \geq N$ one has

$$\|x_n - x_{n-1}\|^2 \geq \frac{1}{S_n} \|\nabla H(y_n, u_n)\|^2 - \frac{2}{s^2 S_n} \|x_{n+1} - x_n\|^2 \geq$$

$$\frac{1}{S_n} \|\nabla H(y_n, u_n)\|^2 - \frac{2}{s^2 S_n \Delta_{n+1}} (r_{n+1} - r_{n+2}).$$

Now by using the Łojasiewicz property of H at $(\overline{x}, \overline{x}) \in \operatorname{crit} H$, and the fact that $\lim_{n \to +\infty} (y_n, u_n) = (\overline{x}, \overline{x})$, we obtain that there exists $K, \epsilon > 0$ and $\overline{N}_1 \in \mathbb{N}$, such that for all $n \geq \overline{N}_1$ one has

$$\|(y_n, u_n) - (\overline{x}, \overline{x})\| < \epsilon,$$

consequently

$$r_n - r_{n+1} \geq \frac{\Delta_n}{S_n} \|\nabla H(y_n, u_n)\|^2 - \frac{2\Delta_n}{s^2 S_n \Delta_{n+1}} (r_{n+1} - r_{n+2}) \geq \tag{22}$$

$$\frac{\Delta_n}{K^2 S_n} r_n^{2\theta} - \frac{2\Delta_n}{s^2 S_n \Delta_{n+1}} (r_{n+1} - r_{n+2}) \geq$$

$$\frac{\Delta_n}{K^2 S_n} r_{n+1}^{2\theta} - \frac{2\Delta_n}{s^2 S_n \Delta_{n+1}} (r_{n+1} - r_{n+2}) = \alpha_n r_{n+1}^{2\theta} - \beta_n (r_{n+1} - r_{n+2}),$$

where $\alpha_n = \frac{\Delta_n}{K^2 S_n}$ and $\beta_n = \frac{2\Delta_n}{s^2 S_n \Delta_{n+1}}$.

It is obvious that the sequences $(\alpha_n)_{n \geq \overline{N}_1}$ and $(\beta_n)_{n \geq \overline{N}_1}$ are convergent, further

$$\lim_{n \to +\infty} \alpha_n > 0 \text{ and } \lim_{n \to +\infty} \beta_n > 0.$$

Now, since $0 < 2\theta \leq 1$ and $r_{n+1} \longrightarrow 0$, there exists $\overline{N}_2 \in \mathbb{N}$, $\overline{N}_2 \geq \overline{N}_1$, such that $r_{n+1}^{2\theta} \geq r_{n+1}$ for all $n \geq \overline{N}_2$.

(∗) Note that this implies that $0 \leq r_n \leq 1$ for all $n \geq \overline{N}_2$.

Hence,

$$r_n \geq (\alpha_n - \beta_n + 1) r_{n+1} + \beta_n r_{n+2}, \text{ for all } n \geq \overline{N}_2.$$

Let us define for every $n > \overline{N}_2$ the sequence $\Xi_n = \frac{\beta_n n^p}{(n+1)^p - n^p}$. Then, since $p > 0$ one has $\lim_{n \to +\infty} \Xi_n = +\infty$.

Since

$$\lim_{n \to +\infty} (\alpha_n - \beta_n + 1) - \left(1 + \frac{\beta_{n+1}}{\Xi_{n+1}} - \frac{\beta_{n-1}}{1 + \frac{\beta_n}{\Xi_n}} \right) = \lim_{n \to +\infty} \alpha_n > 0,$$

there exists $\overline{k} \in \mathbb{N}$, $\overline{k} \geq \overline{N}_2$ such that for all $n \geq \overline{k}$ one has

$$\alpha_n - \beta_n + 1 \geq 1 + \frac{\beta_{n+1}}{\Xi_{n+1}} - \frac{\beta_{n-1}}{1 + \frac{\beta_n}{\Xi_n}}.$$

Consequently,

$$r_n \geq \left(1 + \frac{\beta_{n+1}}{\Xi_{n+1}} - \frac{\beta_{n-1}}{1 + \frac{\beta_n}{\Xi_n}} \right) r_{n+1} + \beta_n r_{n+2}, \text{ for all } n \geq \overline{k},$$

or, equivalently

$$r_n + \frac{\beta_{n-1}}{1 + \frac{\beta_n}{\Xi_n}} r_{n+1} \geq \left(1 + \frac{\beta_{n+1}}{\Xi_{n+1}} \right) \left(r_{n+1} + \frac{\beta_n}{1 + \frac{\beta_{n+1}}{\Xi_{n+1}}} r_{n+2} \right), \tag{23}$$

15

for all $n \geq \overline{k}$. Now (23) leads to

$$\prod_{k=\overline{k}}^{n} \left(r_k + \frac{\beta_{k-1}}{1 + \frac{\beta_k}{\Xi_k}} r_{k+1} \right) \geq \prod_{k=\overline{k}}^{n} \left(1 + \frac{\beta_{k+1}}{\Xi_{k+1}} \right) \prod_{k=\overline{k}}^{n} \left(r_{k+1} + \frac{\beta_k}{1 + \frac{\beta_{k+1}}{\Xi_{k+1}}} r_{k+2} \right),$$

hence after simplifying we get

$$\left(r_{\overline{k}} + \frac{\beta_{\overline{k}-1}}{1 + \frac{\beta_{\overline{k}}}{\Xi_{\overline{k}}}} r_{\overline{k}+1} \right) \prod_{k=\overline{k}}^{n} \frac{1}{1 + \frac{\beta_{k+1}}{\Xi_{k+1}}} \geq r_{n+1} + \frac{\beta_n}{1 + \frac{\beta_{n+1}}{\Xi_{n+1}}} r_{n+2}. \tag{24}$$

But, $\frac{\beta_{n+1}}{\Xi_{n+1}} = \frac{(n+2)^p}{(n+1)^p} - 1$, hence

$$\prod_{k=\overline{k}}^{n} \frac{1}{1 + \frac{\beta_{k+1}}{\Xi_{k+1}}} = \prod_{k=\overline{k}}^{n} \frac{(k+1)^p}{(k+2)^p} = \frac{(\overline{k}+1)^p}{(n+2)^p}.$$

By denoting $\left(r_{\overline{k}} + \frac{\beta_{\overline{k}-1}}{1 + \frac{\beta_{\overline{k}}}{\Xi_{\overline{k}}}} r_{\overline{k}+1} \right) (\overline{k}+1)^p = a_1$, we have

$$a_1 \frac{1}{(n+2)^p} \geq r_{n+1} + \frac{\beta_n}{1 + \frac{1}{n+1}\beta_{n+1}} r_{n+2}.$$

Hence,

$$a_1 \frac{1}{n^p} \geq a_1 \frac{1}{(n+1)^p} \geq r_n = g(y_n) - g(\overline{x}) + \delta_n \|x_n - x_{n-1}\|^2 \geq g(y_n) - g(\overline{x}) \tag{25}$$

which is (a_1).

For (a_2) we start from Lemma 1 and (2) and we have

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

$$\frac{1}{s} \left\langle (x_n - x_{n+1}) + \frac{\beta n}{n + \alpha}(x_n - x_{n-1}), -\frac{\beta n}{n + \alpha}(x_n - x_{n-1}) \right\rangle + \frac{L_g}{2} \left(\frac{\beta n}{n + \alpha} \right)^2 \|x_n - x_{n-1}\|^2 =$$

$$-\left(\frac{\beta n}{n + \alpha} \right)^2 \frac{2 - sL_g}{2s} \|x_n - x_{n-1}\|^2 + \frac{1}{s} \left\langle x_{n+1} - x_n, \frac{\beta n}{n + \alpha}(x_n - x_{n-1}) \right\rangle.$$

By using the inequality $\langle X, Y \rangle \leq \frac{1}{2} \left(a^2 \|X\|^2 + \frac{1}{a^2} \|Y\|^2 \right)$ for all $X, Y \in \mathbb{R}^m, a \in \mathbb{R} \setminus \{0\}$, we obtain

$$\left\langle x_{n+1} - x_n, \frac{\beta n}{n + \alpha}(x_n - x_{n-1}) \right\rangle \leq \frac{1}{2} \left(\frac{1}{2 - sL_g} \|x_{n+1} - x_n\|^2 + (2 - sL_g) \left(\frac{\beta n}{n + \alpha} \right)^2 \|x_n - x_{n-1}\|^2 \right),$$

consequently

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

From (13) we have

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

16

and since the sequence $(g(y_n) + \delta_n \|x_{n+1} - x_n\|^2)_{n \geq \overline{k}}$ is decreasing and has the limit $g(\overline{x})$, we obtain that $g(y_{n+1}) + \delta_{n+1} \|x_{n+1} - x_n\|^2 \geq g(\overline{x})$, consequently

$$\|x_n - x_{n-1}\|^2 \leq \frac{1}{D} r_n. \tag{26}$$

Hence, for all $n \geq \overline{k}$ one has

$$g(x_n) - g(y_n) \leq \frac{1}{2sD(2 - sL_g)} r_{n+1}. \tag{27}$$

Now, the identity $g(x_n) - g(\overline{x}) = (g(x_n) - g(y_n)) + (g(y_n) - g(\overline{x}))$ and (a_1) lead to

$$g(x_n) - g(\overline{x}) \leq \frac{1}{2sD(2 - sL_g)} r_{n+1} + a_1 \frac{1}{n^p}$$

for every $n > \overline{k}$, which combined with (25) give

$$g(x_n) - g(\overline{x}) \leq \frac{1}{2sD(2 - sL_g)} a_1 \frac{1}{(n+2)^p} + a_1 \frac{1}{n^p} \leq a_1 \left(1 + \frac{1}{2sD(2 - sL_g)}\right) \frac{1}{n^p} = a_2 \frac{1}{n^p},$$

for every $n > \overline{k}$.

For (a_3) observe, that by summing up (21) from $n \geq \overline{k}$ to $P > n$ and using the triangle inequality we obtain

$$\|x_P - x_{n-1}\| \leq \sum_{k=n}^{P} \|x_k - x_{k-1}\| \leq$$

$$-\|x_n - x_{n-1}\| + \|x_{P+1} - x_P\| + \frac{9M}{4D}(\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})) - \varphi(H(y_{P+1}, u_{P+1}) - H(\overline{x}, \overline{x}))).$$

By letting $P \longrightarrow +\infty$ we get

$$\|x_{n-1} - \overline{x}\| \leq -\|x_n - x_{n-1}\| + \frac{9M}{4D}\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})) \leq \frac{9M}{4D}\varphi(H(y_n, u_n) - H(\overline{x}, \overline{x})).$$

But, $\varphi(t) = \frac{K}{1-\theta} t^{1-\theta}$, hence

$$\|x_{n-1} - \overline{x}\| \leq \frac{9MK}{4D(1-\theta)}(H(y_n, u_n) - H(\overline{x}, \overline{x}))^{1-\theta} = M_1 r_n^{1-\theta}, \tag{28}$$

where $M_1 = \frac{9MK}{4D(1-\theta)}$.

But $(*)$ assures that $0 \leq r_n \leq 1$ which combined with $\theta \in \left(0, \frac{1}{2}\right]$ leads to $r_n^{1-\theta} \leq \sqrt{r_n}$, consequently we have

$$\|x_{n-1} - \overline{x}\| \leq M_1 \sqrt{r_n}.$$

The conclusion follow by (25), since we have

$$\|x_n - \overline{x}\| \leq M_1 \sqrt{a_1} \frac{1}{n^{\frac{p}{2}}} = a_3 \frac{1}{n^{\frac{p}{2}}}$$

for every $n > \overline{k}$.

Finally, for $n > \overline{k}$ we have

$$\|y_n - \overline{x}\| = \left\|x_n + \frac{\beta n}{n + \alpha}(x_n - x_{n-1}) - \overline{x}\right\| \leq$$

17

$$\left(1 + \frac{\beta n}{n+\alpha}\right) \|x_n - \overline{x}\| + \frac{\beta n}{n+\alpha} \|x_{n-1} - \overline{x}\| \le$$

$$\left(1 + \frac{\beta n}{n+\alpha}\right) a_3 \frac{1}{n^{\frac{p}{2}}} + \frac{\beta n}{n+\alpha} a_3 \frac{1}{n^{\frac{p}{2}}} \le$$

$$\left(1 + 2\frac{\beta n}{n+\alpha}\right) a_3 \frac{1}{n^{\frac{p}{2}}}.$$

Let $a_4 = (1 + 2\beta)a_3$. Then

$$\|y_n - \overline{x}\| \le a_4 \frac{1}{n^{\frac{p}{2}}},$$

for all $n > \overline{k}$, which proves (a_4). ∎

Remark 12 In the previous theorem we obtained convergence rates with order p, for every $p > 0$. This happened when we took in (24)

$$\frac{\beta_{n+1}}{\Xi_{n+1}} = \frac{(n+2)^p}{(n+1)^p} - 1.$$

But actually we have shown more. If one takes $\frac{\beta_{n+1}}{\Xi_{n+1}} = \rho_{n+1} > 0$ where $\lim_{n \to +\infty} \rho_n = 0$ then one obtains that there exits $\overline{k} \in \mathbb{N}$ and $A_1 > 0$ such that for all $n \ge \overline{k}$ one has

$$\alpha_n - \beta_n + 1 \ge 1 + \rho_{n+1} - \frac{\beta_{n-1}}{1 + \rho_n}$$

hence (24) becomes

$$g(y_n) - g(\overline{x}) \le A_1 \prod_{k=\overline{k}+1}^{n} \frac{1}{1 + \rho_k}.$$

From here, as in the proof of Theorem 11, one can derive that

$$g(x_n) - g(\overline{x}) \le A_2 \prod_{k=\overline{k}+1}^{n} \frac{1}{1 + \rho_k} \text{ for some } A_2 > 0,$$

and

$$\|x_n - \overline{x}\| = \mathcal{O}\left(\sqrt{\prod_{k=\overline{k}+1}^{n} \frac{1}{1 + \rho_k}}\right) \text{ and } \|y_n - \overline{x}\| = \mathcal{O}\left(\sqrt{\prod_{k=\overline{k}+1}^{n} \frac{1}{1 + \rho_k}}\right).$$

Having in mind this general result, and taking into account that in [14], for the dynamical system (6) which, as it is shown in Introduction, can be viewed as the continuous counterpart of the numerical scheme (2), it was obtained finite time convergence of the generated trajectories for $\theta \in \left(0, \frac{1}{2}\right)$ and exponential convergence rate for $\theta = \frac{1}{2}$, it seems a valid question whether we can obtain exponential convergence rate for the sequences generated by (2), by choosing an appropriate sequence ρ_n. We show in what follow that this is not possible. We have

$$\prod_{k=\overline{k}+1}^{n} \frac{1}{1 + \rho_k} = e^{-\sum_{k=\overline{k}+1}^{n} \ln(1+\rho_k)}.$$

Obviously $\ln(1 + \rho_k) > 0$, for all $k > \overline{k}$ and $\lim_{k \to +\infty} \ln(1 + \rho_k) = 0$. Now, by using the Cesàro-Stolz theorem we obtain that

$$\lim_{n \to +\infty} \frac{\sum_{k=\overline{k}+1}^{n} \ln(1 + \rho_k)}{n} = \lim_{n \to +\infty} \ln(1 + \rho_{n+1}) = 0,$$

18

hence $\sum_{k=\bar{k}+1}^{n} \ln(1 + \rho_k) = \mathfrak{o}(n)$, which shows that

$$\mathcal{O}\left(\prod_{k=\bar{k}+1}^{n} \frac{1}{1 + \rho_k}\right) > \mathcal{O}\left(e^{-n}\right).$$

Remark 13 According to [18], H is KL with Łojasiewicz exponent $\theta \in \left[\frac{1}{2}, 1\right)$, whenever g is KL with Łojasiewicz exponent $\theta \in \left[\frac{1}{2}, 1\right)$. Therefore, we have the following corollary.

Corollary 14 *In the settings of problem* (1) *consider the sequences* $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ *generated by Algorithm* (2). *Assume that* g *is bounded from below and that* $(x_n)_{n\in\mathbb{N}}$ *is bounded, let* $\overline{x} \in \mathrm{crit}(g)$ *be such that* $\lim_{n\longrightarrow+\infty} x_n = \overline{x}$ *and suppose that* g *fulfills the Łojasiewicz property at* $\overline{x}$ *with Łojasiewicz exponent* $\theta = \frac{1}{2}$. *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$_1$) $g(y_n) - g(\overline{x}) \leq a_1 \frac{1}{n^p}$ *for every* $n > \bar{k}$,

(a$_2$) $g(x_n) - g(\overline{x}) \leq a_2 \frac{1}{n^p}$ *for every* $n > \bar{k}$,

(a$_3$) $\|x_n - \overline{x}\| \leq a_3 \frac{1}{n^{\frac{p}{2}}}$ *for every* $n > \bar{k}$,

(a$_4$) $\|y_n - \overline{x}\| \leq a_4 \frac{1}{n^{\frac{p}{2}}}$ *for all* $n > \bar{k}$.

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

Theorem 15 *In the settings of problem* (1) *consider the sequences* $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ *generated by Algorithm* (2). *Assume that* g *is bounded from below and that* $(x_n)_{n\in\mathbb{N}}$ *is bounded, let* $\overline{x} \in \mathrm{crit}(g)$ *be such that* $\lim_{n\longrightarrow+\infty} x_n = \overline{x}$ *and suppose that*

$$H : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}, \ H(x,y) = g(x) + \frac{1}{2}\|x - y\|^2$$

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

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

(b$_2$) $g(x_n) - g(\overline{x}) \leq b_2 \frac{1}{n^{\frac{1}{2\theta-1}}}$, *for all* $n \geq \overline{N}_1 + 2$;

(b$_3$) $\|x_n - \overline{x}\| \leq b_3 \frac{1}{n^{\frac{1-\theta}{2\theta-1}}}$, *for all* $n \geq \overline{N}_1 + 2$;

(b$_4$) $\|y_n - \overline{x}\| \leq b_4 \frac{1}{n^{\frac{1-\theta}{2\theta-1}}}$, *for all* $n > \overline{N}_1 + 2$,

where $\overline{N}_1 \in \mathbb{N}$ *was defined in the proof of Theorem 11 .*

Proof. Also here, to avoid triviality, in what follows we assume that $H(y_n, u_n) > H(\overline{x}, \overline{x})$, for all $n \geq N$. From (22) we have that for every $n \geq \overline{N}_1$ it holds

$$r_n - r_{n+1} \geq \alpha_n r_{n+1}^{2\theta} - \beta_n(r_{n+1} - r_{n+2}),$$

where $\alpha_n = \frac{\Delta_n}{K^2 S_n}$ and $\beta_n = \frac{2\Delta_n}{s^2 S_n \Delta_{n+1}}$.

19

Hence,
$$(r_n - r_{n+1})r_{n+1}^{-2\theta} + \beta_n(r_{n+1} - r_{n+2})r_{n+1}^{-2\theta} \geq \alpha_n,$$

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

Consider the function $\phi(t) = \frac{K}{2\theta-1}t^{1-2\theta}$ where K is the constant defined at the Łojasiewicz property of H. Then $\phi'(t) = -Kt^{-2\theta}$ and we have

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

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

Assume that for some $n \geq \overline{N}_1$ it holds that $r_n^{-2\theta} \geq \frac{1}{2}r_{n+1}^{-2\theta}$.

Then

$$\phi(r_{n+1}) - \phi(r_n) + \beta_n(\phi(r_{n+2}) - \phi(r_{n+1})) \geq \frac{K}{2}(r_n - r_{n+1})r_{n+1}^{-2\theta} + K\beta_n(r_{n+1} - r_{n+2})r_{n+1}^{-2\theta} \geq \quad (29)$$

$$\frac{K}{2}(r_n - r_{n+1})r_{n+1}^{-2\theta} + \frac{K}{2}\beta_n(r_{n+1} - r_{n+2})r_{n+1}^{-2\theta} \geq \frac{K}{2}\alpha_n.$$

Conversely, if $2r_n^{-2\theta} < r_{n+1}^{-2\theta}$ for some $n \geq \overline{N}_1$, then

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

hence,

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

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

Let $C_2 = \inf_{n \geq \overline{N}_1} \frac{C_1(1+\beta_n)}{\alpha_n} > 0$. Then,

$$\phi(r_{n+1}) - \phi(r_n) + \beta_n(\phi(r_{n+2}) - \phi(r_{n+1})) \geq C_2\alpha_n. \quad (30)$$

From (29) and (30) we get that there exists $C > 0$ such that

$$\phi(r_{n+1}) - \phi(r_n) + \beta_n(\phi(r_{n+2}) - \phi(r_{n+1})) \geq C\alpha_n, \text{ for all } n \geq \overline{N}_1.$$

Let $\overline{\beta} = \sup_{n \geq \overline{N}_1} \beta_n$. Then the latter relation becomes

$$\phi(r_{n+1}) - \phi(r_n) + \overline{\beta}(\phi(r_{n+2}) - \phi(r_{n+1})) \geq C\alpha_n, \text{ for all } n \geq \overline{N}_1,$$

which leads to

$$\sum_{k=\overline{N}_1}^{n}\left(\phi(r_{k+1}) - \phi(r_k) + \overline{\beta}(\phi(r_{k+2}) - \phi(r_{k+1}))\right) \geq C\sum_{k=\overline{N}_1}^{n}\alpha_k.$$

Consequently,

$$\phi(r_{n+1}) - \phi(r_{\overline{N}_1}) + \overline{\beta}(\phi(r_{n+2}) - \phi(r_{\overline{N}_1+1})) \geq C\sum_{k=\overline{N}_1}^{n}\alpha_k$$

and by using the fact that the sequence $(r_n)_{n \geq \overline{N}_1}$ is decreasing and ϕ is also decreasing, we obtain

$$(1 + \overline{\beta})\phi(r_{n+2}) \geq C \sum_{k=\overline{N}_1}^{n} \alpha_k.$$

In other words

$$r_n^{1-2\theta} \geq \frac{C(2\theta - 1)}{K(1 + \overline{\beta})} \sum_{k=\overline{N}_1}^{n-2} \alpha_k, \text{ for all } n \geq \overline{N}_1 + 2.$$

Hence,

$$r_n \leq \left(\frac{C(2\theta - 1)}{K(1 + \overline{\beta})} \right)^{\frac{-1}{2\theta - 1}} \left(\sum_{k=\overline{N}_1}^{n-2} \alpha_k \right)^{\frac{-1}{2\theta - 1}}, \text{ for all } n \geq \overline{N}_1 + 2.$$

Since $\sum_{k=\overline{N}_1}^{n-2} \alpha_k \geq \underline{\alpha}(n - \overline{N}_1 - 1)$, where $0 < \underline{\alpha} = \inf_{k \geq \overline{N}_1} \alpha_k$ we have that there exists $M > 0$ such that

$$\left(\sum_{k=\overline{N}_1}^{n-2} \alpha_k \right)^{\frac{-1}{2\theta - 1}} \leq \underline{\alpha}^{\frac{-1}{2\theta - 1}} (n - \overline{N}_1 - 1)^{\frac{-1}{2\theta - 1}} \leq \underline{\alpha}^{\frac{-1}{2\theta - 1}} M n^{\frac{-1}{2\theta - 1}}, \text{ for all } n \geq \overline{N}_1 + 2.$$

Therefore, we have

$$r_n \leq \left(\frac{C(2\theta - 1)}{K(1 + \overline{\beta})} \right)^{\frac{-1}{2\theta - 1}} \underline{\alpha}^{\frac{-1}{2\theta - 1}} M n^{\frac{-1}{2\theta - 1}} = b_1 n^{\frac{-1}{2\theta - 1}}, \text{ for all } n \geq \overline{N}_1 + 2.$$

But, $r_n = g(y_n) - g(\overline{x}) + \delta_n \|x_n - x_{n-1}\|^2$, consequently

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

and (b_1) is proved.

For (b_2) observe that (27) holds for all $n \geq \overline{N}_1$, hence for all $n \geq \overline{N}_1$ one has

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

Thus, there exists $M > 0$ such that

$$g(x_n) - g(\overline{x}) = (g(x_n) - g(y_n)) + (g(y_n) - g(\overline{x})) \leq \left(\frac{1}{2sD(2 - sL_g)} b_1 M + b_1 \right) n^{\frac{-1}{2\theta - 1}} = b_2 n^{\frac{-1}{2\theta - 1}},$$

for all $n \geq \overline{N}_1 + 2$.

For proving (b_3) we use (28). Note that the relation $\|x_n - \overline{x}\| \leq M_1 r_n^{1-\theta}$ holds for all $n \geq \overline{N}_1$. Hence,

$$\|x_n - \overline{x}\| \leq M_1 \left(b_1 n^{\frac{-1}{2\theta - 1}} \right)^{1-\theta}, \text{ for all } n \geq \overline{N}_1 + 2.$$

Consequently,

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

where $b_3 = M_1 b_1^{1-\theta}$ and this proves (b_3).

For (b_4) observe that for $n \geq \overline{N}_1 + 3$ we have

$$\|y_n - \overline{x}\| = \left\| x_n + \frac{\beta n}{n + \alpha}(x_n - x_{n-1}) - \overline{x} \right\| \leq$$

21

$$\left(1 + \frac{\beta n}{n + \alpha}\right)\|x_n - \overline{x}\| + \frac{\beta n}{n + \alpha}\|x_{n-1} - \overline{x}\| \le$$

$$\left(1 + \frac{\beta n}{n + \alpha}\right) b_3 n^{\frac{\theta-1}{2\theta-1}} + \frac{\beta n}{n + \alpha} b_3(n - 1)^{\frac{\theta-1}{2\theta-1}} \le$$

$$\left(\left(1 + \frac{\beta n}{n + \alpha}\right) b_3 + \frac{\beta n}{n + \alpha} b_3\right)(n - 1)^{\frac{\theta-1}{2\theta-1}} \le b_4 n^{\frac{\theta-1}{2\theta-1}},$$

where one can take $b_4 = \sup_{n \ge \overline{N}_1 + 3}\left(\left(1 + \frac{\beta n}{n+\alpha}\right)b_3 + \frac{\beta n}{n+\alpha}b_3\right)\left(\frac{n}{n-1}\right)^{\frac{1-\theta}{2\theta-1}}$. ∎

According to Remark 13 we have the following corollary.

Corollary 16 *In the settings of problem (1) consider the sequences $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$ generated by Algorithm (2). Assume that g is bounded from below and that $(x_n)_{n \in \mathbb{N}}$ is bounded, let $\overline{x} \in \mathrm{crit}(g)$ be such that $\lim_{n \longrightarrow +\infty} x_n = \overline{x}$ and suppose that g fulfills the Łojasiewicz property at $\overline{x}$ with Łojasiewicz exponent $\theta \in \left(\frac{1}{2}, 1\right)$. Then, there exist $b_1, b_2, b_3, b_4 > 0$ such that the following statements hold true:*

(b₁) $g(y_n) - g(\overline{x}) \le b_1 \frac{1}{n^{\frac{1}{2\theta-1}}}$, *for all $n \ge \overline{N}_1 + 2$;*

(b₂) $g(x_n) - g(\overline{x}) \le b_2 \frac{1}{n^{\frac{1}{2\theta-1}}}$, *for all $n \ge \overline{N}_1 + 2$;*

(b₃) $\|x_n - \overline{x}\| \le b_3 \frac{1}{n^{\frac{1-\theta}{2\theta-1}}}$, *for all $n \ge \overline{N}_1 + 2$;*

(b₄) $\|y_n - \overline{x}\| \le b_4 \frac{1}{n^{\frac{1-\theta}{2\theta-1}}}$, *for all $n > \overline{N}_1 + 2$,*

where $\overline{N}_1 \in \mathbb{N}$ was defined in the proof of Theorem 11.

4   Conclusions

In this paper we show the convergence of a Nesterov type algorithm in a full nonconvex setting, assuming that a regularization of the objective function satisfies the Kurdyka-Łojasiewicz property. For this purpose as a starting point we show a sufficient decrease property for the iterates generated by our algorithm. Though our algorithm is asymptotically equivalent to Nesterov's accelerated gradient method, we cannot obtain full equivalence due to the fact that in order to obtain the above mentioned decrease property we cannot allow the inertial parameter, more precisely the parameter β, to attain the value 1. Nevertheless, we obtain convergence rates of order p for every $p > 0$, for the sequences generated by our numerical scheme but also for the function values in these sequences, provided the objective function, or a regularization of the objective function, satisfies the Łojasiewicz property with Łojasiewicz exponent $\theta \in \left(0, \frac{1}{2}\right]$. We also show that, at least with our techniques, exponential convergence rates cannot be obtained. In case the Łojasiewicz exponent of the objective function, or a regularization of the objective function, is $\theta \in \left(\frac{1}{2}, 1\right)$, we obtain polynomial convergence rates.

A related future research is the study of a modified FISTA algorithm in a nonconvex setting. Indeed, let $f : \mathbb{R}^m \longrightarrow \overline{\mathbb{R}}$ be a proper convex and lower semicontinuous function and let $g : \mathbb{R}^m \longrightarrow \mathbb{R}$ be a (possible nonconvex) smooth function with L_g Lipschitz continuous gradient. Consider the optimization problem

$$\inf_{x \in \mathbb{R}^m} f(x) + g(x).$$

We associate to this optimization problem the following proximal-gradient algorithm. For $x_0, y_0 \in \mathbb{R}^m$ consider

$$\begin{cases} x_{n+1} = \mathrm{prox}_{sf}(y_n - s\nabla g(y_n)), \\ \\ y_n = x_n + \dfrac{\beta n}{n + \alpha}(x_n - x_{n-1}), \end{cases} \tag{31}$$

where $\alpha > 0$, $\beta \in (0,1)$ and $0 < s < \frac{2(1-\beta)}{L_g}$. Obviously, when $f \equiv 0$ then (31) becomes the numerical scheme (2) studied in the present paper.

We emphasize that (31) has a similar formulation as the modified FISTA algorithm studied by Chambolle and Dossal in [15] and the convergence of the generated sequences, to a critical point of the objective function $f + g$, would open the gate for the study of FISTA type algorithms in a nonconvex setting.

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-Łojasiewicz 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] H. Attouch, Z. Chbani, H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$*, ESAIM: COCV (2017). doi:10.1051/cocv/2017083

[6] H. Attouch, J. Peypouquet, P. Redont, *Fast convex optimization via inertial dynamics with Hessian driven damping*, J. Differential Equations 261(10), 5734-5783, 2016

[7] A. Beck, M. Teboulle, *A Fast Iterative Shrinkage- Thresholding Algorithm for Linear Inverse Problems*, SIAM Journal on Imaging Sciences 2(1), 183-202, 2009

[8] 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

[9] J. Bolte, A. Daniilidis, A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization 17(4), 1205-1223, 2006

[10] J. Bolte, A. Daniilidis, A. Lewis, M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization 18(2), 556-572, 2007

[11] J. Bolte, A. Daniilidis, O. Ley, L. Mazet, *Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society 362(6), 3319-3363, 2010

[12] R.I. Boţ, E.R. Csetnek, S.C. László, *Approaching nonsmooth nonconvex minimization through second-order proximal-gradient dynamical systems*, Journal of Evolution Equations (2018). https://doi.org/10.1007/s00028-018-0441-7

[13] R.I. Boţ, 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

[14] R.I. Boţ, 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

[15] A. Chambolle, Ch. Dossal, *On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm"*, J. Optim. Theory Appl. 166(3), 968-982, 2015

[16] P. Frankel, G. Garrigos, J. Peypouquet, *Splitting Methods with Variable Metric for KurdykaŁojasiewicz Functions and General Convergence Rates*, Journal of Optimization Theory and Applications, 165(3), 874900, 2015

[17] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Annales de l'institut Fourier (Grenoble) 48(3), 769-783, 1998

[18] G. Li, T. K. Pong, *Calculus of the Exponent of Kurdyka-Łojasiewicz Inequality and Its Applications to Linear Convergence of First-Order Methods*, Foundations of Computational Mathematics, 1-34 , 2018

[19] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique Paris, 87-89, 1963

[20] D.A. Lorenz, T. Pock, *An inertial forwardbackward algorithm for monotone inclusions*, J. Math. Imaging Vis., 51(2), 311-325 2015

[21] 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

[22] Y. Nesterov , *Introductory lectures on convex optimization: a basic course*. Kluwer Academic Publishers, Dordrecht, 2004

[23] 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

[24] R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis*, Fundamental Principles of Mathematical Sciences 317, Springer-Verlag, Berlin, 1998

[25] 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