Approaching nonsmooth nonconvex minimization through second order proximal-gradient dynamical systems

Szilárd László

Department of Mathematics
Technical University of Cluj-Napoca

GAMES, DYNAMICS and OPTIMIZATION
March 13-15, 2018
Wien, Austria

Consider the optimization problem

$$\inf_{x\in\mathbb{R}^n}(f(x)+g(x)), \tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a proper, convex and lower semicontinuous function and
$g:\mathbb{R}^n\to\mathbb{R}$ is a (possibly nonconvex) Fréchet differentiable function with β-Lipschitz continuous gradient, i.e. there exists $\beta\geq 0$ such that

$$\|\nabla g(x)-\nabla g(y)\|\leq\beta\|x-y\| \text{ for all } x,y\in\mathbb{R}^n.$$

We associate to (1) the following second order dynamical system of implicit-type

$$\left\{ \begin{array}{l} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) = \operatorname{prox}_{\lambda f}\left(x(t) - \lambda\nabla g(x(t))\right) \\ x(0) = u_0,\ \dot{x}(0) = v_0, \end{array} \right. \tag{2}$$

where $u_0, v_0 \in \mathbb{R}^n$, $\gamma, \lambda \in (0, +\infty)$ and

$$\operatorname{prox}_{\lambda f} : \mathbb{R}^n \to \mathbb{R}^n, \quad \operatorname{prox}_{\lambda f}(x) = \operatorname*{argmin}_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2\lambda}\|y - x\|^2 \right\},$$

denotes the proximal point operator of λf.

Dynamical systems of proximal-gradient type associated to optimization problems have been intensively treated in the literature. Bolte (2003), studied the convergence of the trajectories of the first order dynamical system

$$\begin{cases} \dot{x}(t) + x(t) = \text{proj}_C \left(x(t) - \lambda \nabla g(x(t))\right) \\ x(0) = x_0, \end{cases}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a convex smooth function, $C \subseteq \mathbb{R}^n$ is a nonempty, closed and convex set, $x_0 \in \mathbb{R}^n$, and proj_C denotes the projection operator on the set C. The trajectory of this system has been proved to converge to a minimizer of the optimization problem

$$\inf_{x \in C} g(x),$$

provided the latter is solvable.

The following first order dynamical system

$$\left\{ \begin{array}{l} \dot{x}(t) + x(t) = \text{prox}_{\lambda f}\left(x(t) - \lambda \nabla g(x(t))\right) \\ x(0) = x_0, \end{array} \right.$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function, $g : \mathbb{R}^n \to \mathbb{R}$ is a convex smooth function and $x_0 \in \mathbb{R}^n$, has been recently considered by Abbas and Attouch (2015) and, in case g is possible nonconvex, by Boț and Csetnek (2018), in relation to the optimization problem

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

In case this problem is solvable, the trajectory generated by the dynamical system has been proved to converge to a global minimizer of it.

The second order projected-gradient system

$$\begin{cases} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) = \text{proj}_C(x(t) - \lambda\nabla g(x(t))) \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases}$$

with damping parameter $\gamma > 0$ and step size $\lambda > 0$, has been considered by Antipin (1994), and Attouch and Alvarez (1998), in connection with the problem

$$\inf_{x \in \mathbb{R}^n} g(x).$$

Existence and uniqueness

By making use of the notation $X(t) = (x(t), \dot{x}(t))$, the system (2) can be rewritten as

$$\begin{cases} \dot{X}(t) = F(X(t)) \\ X(0) = (u_0, v_0), \end{cases} \tag{3}$$

where $F : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n \times \mathbb{R}^n$,

$$F(u, v) = \left(v, \operatorname*{prox}_{\lambda f} \left(u - \lambda \nabla g(u)\right) - \gamma v - u \right).$$

We prove the existence and uniqueness of a global solution of (3) by using the Cauchy-Lipschitz Theorem. To this aim it is enough to show that F is globally Lipschitz continuous.

Existence and uniqueness

The following holds.

Theorem

For every starting points $u_0, v_0 \in \mathbb{R}^n$, the dynamical system (2) has a unique global solution $x \in C^2([0,+\infty), \mathbb{R}^n)$. Further, $\dddot{X}$ exists almost everywhere on $[0,+\infty)$ and for almost every $t \in [0,+\infty)$ one has

$$\|\ddot{X}(t)\| \leq L\|\dot{X}(t)\|,$$

where L is the Lipschitz constant of F.

Consequently, for a trajectory x of (2), for almost every $t \in [0,+\infty)$ we have the following estimate of the third order derivative

$$\|x^{(3)}(t)\|^2 \leq L^2\|\dot{x}(t)\|^2 + (L^2-1)\|\ddot{x}(t)\|^2$$

We have the following Lipschitz constant for F.

$$L_1 := \sqrt{\max\left((\gamma+1)^2, (\gamma+2)((1+\lambda\beta)^2+1)\right)}$$

Consequently, by using L_1 we have that for almost every $t \in [0, +\infty)$

$$\|x^{(3)}(t)\|^2 \le \max\left((\gamma+1)^2, (\gamma+2)((1+\lambda\beta)^2+1)\right)\|\dot{x}(t)\|^2 +$$

$$(\max\left((\gamma+1)^2, (\gamma+2)((1+\lambda\beta)^2+1)\right) - 1)\|\ddot{x}(t)\|^2.$$

Another Lipschitz constant is

$$L_2 := \sqrt{\max((\gamma+1)^2+\gamma\lambda\beta, (2+\lambda\beta)^2+\gamma(2+\lambda\beta))}.$$

By using L_2, one obtains for almost every $t \in [0, +\infty)$

$$\|x^{(3)}(t)\|^2 \leq \max((\gamma+1)^2+\gamma\lambda\beta, (2+\lambda\beta)^2+\gamma(2+\lambda\beta))\|\dot{x}(t)\|^2 +$$

$$(\max((\gamma+1)^2+\gamma\lambda\beta, (2+\lambda\beta)^2+\gamma(2+\lambda\beta)) - 1)\|\ddot{x}(t)\|^2.$$

Existence and uniqueness

Remark

Obviously, $L_1 > 2$ and $L_2 > 2$. One can easily verify that $L_2 \leq L_1$, provided $\gamma \leq \sqrt{3}$. Moreover, if $\gamma \leq \sqrt{3}$, then

$$L_2 = \sqrt{(2+\lambda\beta)^2 + \gamma(2+\lambda\beta)}.$$

However, for $\gamma > \sqrt{3}$, one may have $L_2 > L_1$ and also $L_2 < L_1$. Indeed, for $\gamma = 2$ and $\lambda\beta = \frac{1}{10}$, it holds

$$L_2 = \sqrt{9,2} > 3 = L_1,$$

while for $\gamma = 2$ and $\lambda\beta = 1$ it holds

$$L_2 = \sqrt{15} < \sqrt{20} = L_1.$$

Let $T > 0$ and let $x \in C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (2). Then, for almost every $t \in [0, T]$ we have

$$\frac{d}{dt}\Big[(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))+ \frac{1}{2\lambda}\|\ddot{x}(t)+c\gamma\dot{x}(t)\|^2 - \frac{C}{2\lambda}\|\dot{x}(t)\|^2\Big] \leq A\|\dot{x}(t)\|^2 + B\|\ddot{x}(t)\|^2,$$

where $c = \frac{L^2}{L^2+1}$, $L = \min(L_1, L_2)$ further

$$A = -\frac{1}{2}\frac{\gamma}{\lambda} + \frac{\beta}{2}(L^2 + 2\gamma^2 + 1),$$

$$B = -\frac{1}{2L^2}\frac{\gamma}{\lambda} + \frac{\beta}{2}(L^2 + \gamma^2 + 1)$$

and

$$C = -\frac{(2L^2+1)}{(L^2+1)^2}\gamma^2 + 3\beta\gamma\lambda - 1.$$

Consequently the following result holds.

Lemma

Suppose that $f+g$ *is bounded from bellow and* $\gamma, \lambda > 0$ *satisfy the following set of conditions:*

$$(\rho) \quad A < 0,\ B < 0,\ \text{and } C < 0.$$

For $u_0, v_0 \in \mathbb{R}^n$*, let* $x \in C^2([0,+\infty), \mathbb{R}^n)$ *be the unique global solution of* (2)*. Then the following statements are true*

(a) $\dot{x} \in L^2([0,+\infty), \mathbb{R}^n)$ *and* $\lim_{t \longrightarrow +\infty} \dot{x}(t) = 0$;

(b) $\ddot{x} \in L^2([0,+\infty), \mathbb{R}^n)$ *and* $\lim_{t \longrightarrow +\infty} \ddot{x}(t) = 0$;

(c) $\exists \lim_{t \longrightarrow +\infty} (f+g)(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) \in \mathbb{R}$.

The choice $\gamma\lambda\beta \leq \frac{1}{3}$ guarantees that $C < 0$. Moreover, in this case $B > A$. Thus, we have.

Corollary

Suppose that $f + g$ is bounded from bellow and $\sqrt{3} \geq \gamma > 0, \lambda > 0$ satisfy the following condition

$$-\frac{1}{(2+\lambda\beta)^2 + \gamma(2+\lambda\beta)}\frac{\gamma}{\lambda} + \beta((2+\lambda\beta)^2 + \gamma(2+\lambda\beta) + \gamma^2 + 1) < 0.$$

For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty),\mathbb{R}^n)$ be the unique global solution of (2). Then the following statements are true

(a) $\dot{x} \in L^2([0,+\infty),\mathbb{R}^n)$ *and* $\lim_{t\longrightarrow+\infty} \dot{x}(t) = 0$;

(b) $\ddot{x} \in L^2([0,+\infty),\mathbb{R}^n)$ *and* $\lim_{t\longrightarrow+\infty} \ddot{x}(t) = 0$;

(c) $\exists \lim_{t\longrightarrow+\infty}(f+g)(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) \in \mathbb{R}$.

We present two examples of γ, λ such that the set of conditions (ρ) holds.

Example

Consider $\gamma = 3$ and $\lambda\beta = \frac{1}{200}$.
Then $L = 4$,

$$A = \frac{1}{2\lambda}(-\gamma + \lambda\beta(L^2 + 2\gamma^2 + 1)) = -\frac{1}{2\lambda} \cdot \frac{565}{200} < 0,$$

$$B = \frac{1}{2\lambda L^2}(-\gamma + \lambda\beta L^2(L^2 + \gamma^2 + 1)) = -\frac{1}{4\lambda} \cdot \frac{23}{200} < 0$$

and

$$C = -\frac{(2L^2 + 1)}{(L^2 + 1)^2}\gamma^2 + 3\beta\gamma\lambda - 1 = -\frac{586}{289} + \frac{9}{200} < 0.$$

Example

Consider $\gamma = 1$ and $\lambda\beta = \frac{1}{50}$. According to the previous corollary is enough to show that

$$2B = -\frac{1}{(2+\lambda\beta)^2+\gamma(2+\lambda\beta)}\frac{\gamma}{\lambda}+\beta((2+\lambda\beta)^2+\gamma(2+\lambda\beta)+\gamma^2+1) < 0.$$

We have

$$2B = -\frac{1}{\lambda}\left(\frac{10000}{61004} - \frac{1}{50}\cdot\frac{81004}{10000}\right) =$$

$$-\frac{1}{\lambda}(0,1639... - 0,162008) < 0,$$

hence (ρ) hold.

Lemma

Assume that $f+g$ is bounded from below and γ, λ satisfy the set of conditions (ρ). For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty), \mathbb{R}^n)$ be the unique global solution of (2). Then the set of limit points of x,, which we denote by $\omega(x)$, is a subset of the set of critical points of $f+g$. In other words,

$$\omega(x) := \{\overline{x} \in \mathbb{R}^n : \exists t_k \to \infty \text{ such that } x(t_k) \to \overline{x},\ k \to +\infty\} \subseteq$$

$$\operatorname{crit}(f+g) := \{x \in \mathbb{R}^n : 0 \in \partial(f+g)(x) = \partial f(x) + \nabla g(x)\}.$$

We have seen that for $x \in C^2([0,+\infty),\mathbb{R}^n)$ a unique global solution of (2). Then, for $T > 0$ and almost every $t \in [0, T]$ we have

$$\frac{d}{dt}\Big[(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))+ \frac{1}{2\lambda}\|\ddot{x}(t)+c\gamma\dot{x}(t)\|^2-\frac{C}{2\lambda}\|\dot{x}(t)\|^2\Big] \leq A\|\dot{x}(t)\|^2+B\|\ddot{x}(t)\|^2 \leq 0.$$

This suggest to consider the regularization function

$$H:\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^n \longrightarrow \mathbb{R}\cup\{+\infty\},$$

$$H(u,v,w)=(f+g)(u)+\frac{1}{2\lambda}\|u-v\|^2-\frac{C}{2\lambda}\|w\|^2.$$

Lemma

Assume that $f+g$ *is bounded from below and* γ, λ *satisfy the set of conditions* (ρ). *For* $u_0, v_0 \in \mathbb{R}^n$, *let* $x \in C^2([0,+\infty), \mathbb{R}^n)$ *be the unique global solution of* (2). *Then the following statements are true*

(H_1*) for almost every* $t \in [0,+\infty)$ *it holds*

$$\frac{d}{dt}\left(H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t), \gamma(1-c)\dot{x}(t)+x(t), \dot{x}(t))\right) \leq 0$$

and the limit

$$\lim_{t \longrightarrow +\infty} H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t), \gamma(1-c)\dot{x}(t)+x(t), \dot{x}(t))$$

exists and is finite, where $c = \frac{L^2}{L^2+1}$;

(H$_2$) for almost every $t \in [0, +\infty)$ and for every $a \geq 0$ we have

$$w(t) = \Big(-\nabla g(x(t)) + \nabla g(\ddot{x}(t) + \gamma \dot{x}(t) + x(t)) - \frac{1}{\lambda} a\gamma \dot{x}(t)),$$

$$-\frac{1}{\lambda}(\ddot{x}(t) + (1-a)\gamma \dot{x}(t)), -\frac{C}{\lambda}\dot{x}(t) \Big) \in$$

$$\partial H(\ddot{x}(t) + \gamma \dot{x}(t) + x(t), \gamma a \dot{x}(t) + x(t), \dot{x}(t))$$

and

$$\|w(t)\| \leq \left(\beta + \frac{1}{\lambda}\right) \|\ddot{x}(t)\| + \frac{\beta\lambda\gamma + (2a+1)\gamma - C}{\lambda} \|\dot{x}(t)\|;$$

(H$_3$) for $\overline{x} \in \omega(x)$ and $t_k \longrightarrow +\infty$ such that $x(t_k) \longrightarrow \overline{x}$ as $k \longrightarrow +\infty$, and for every $a \geq 0$ we have

$$H(\ddot{x}(t_k) + \gamma\dot{x}(t_k) + x(t_k), a\gamma\dot{x}(t) + x(t_k), \dot{x}(t_k)) \longrightarrow H(\overline{x}, \overline{x}, 0)$$

as $k \longrightarrow +\infty$.

Lemma

Assume that $f+g$ is bounded from below and γ, λ satisfy the set of conditions (ρ). For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty), \mathbb{R}^n)$ be the unique global solution of (2). Suppose that x is bounded and let $a \geq 0$. Then the following statements are true

(a) $\omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x}) \subseteq \text{crit}(H) = \{(u,u,0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : u \in \text{crit}(f+g)\}$;

(b) $\lim\limits_{t \longrightarrow +\infty} \text{dist}((\ddot{x}(t)+\gamma\dot{x}(t)+x(t), a\gamma\dot{x}(t)+x(t), \dot{x}(t)), \omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})) = 0$;

(c) *H is finite and constant on* $\omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})$;

(d) $\omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})$ *is nonempty, compact and connected.*

The convergence of the trajectory generated by the dynamical system (2) will be shown in the framework of functions satisfying the *Kurdyka-Łojasiewicz property*.
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 $s \in (0, \eta)$.
In the following definition we use the *distance function* to a set, defined for $A \subseteq \mathbb{R}^n$ as $\operatorname{dist}(x, A) = \inf_{y \in A} \|x - y\|$ for all $x \in \mathbb{R}^n$.

Definition

(*Kurdyka-Łojasiewicz property*) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. We say that f satisfies the *Kurdyka-Łojasiewicz (KL) property* at $\overline{x} \in \operatorname{dom} \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$ 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

$$\varphi'(f(x) - f(\overline{x}))\operatorname{dist}(0, \partial f(x)) \geq 1.$$

If f satisfies the KL property at each point in $\operatorname{dom} \partial f$, then f is called a *KL function*.

A main result.

Theorem

Assume that $f+g$ is bounded from below and γ, λ satisfy the set of conditions (ρ). For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty), \mathbb{R}^n)$ be the unique global solution of (2). Suppose that x is bounded and H is a KL function. Then the following statements are true

(a) $\dot{x} \in L^1([0,+\infty), \mathbb{R}^n)$;

(b) $\ddot{x} \in L^1([0,+\infty), \mathbb{R}^n)$;

(c) *there exists* $\overline{x} \in \operatorname{crit}(f+g)$ *such that* $\lim_{t \longrightarrow +\infty} x(t) = \overline{x}$.

Remark

Since the class of semi-algebraic functions is closed under addition and $(u, v) \mapsto \alpha\|u - v\|^2$ and $w \mapsto \alpha'\|w\|^2$ are semi-algebraic for $\alpha, \alpha' > 0$, the conclusion of the previous theorem holds if the condition H is a KL function is replaced by the assumption that $f + g$ is semi-algebraic.

Remark

Assume that $\gamma, \lambda > 0$ fulfill the set of conditions (ρ) and that $f + g$ is coercive, that is

$$\lim_{\|u\| \to +\infty} (f + g)(u) = +\infty.$$

For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (2). Then x is bounded.

In the context of optimization problems involving KL functions, it is known that convergence rates of the trajectory can be formulated in terms of the so-called Łojasiewicz exponent.

Definition

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous 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

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

$$\text{and every } x^* \in \partial f(x).$$

The number θ is called the Łojasiewicz exponent of f at the critical point $\overline{x}$.

In the following theorem we obtain convergence rates for both the trajectory generated (2) and its velocity.

Theorem

Assume that $f+g$ *is bounded from below and* γ, λ *satisfy the set of conditions* (ρ). *For* $u_0, v_0 \in \mathbb{R}^n$, *let* $x \in C^2([0,+\infty), \mathbb{R}^n)$ *be the unique global solution of* (2). *Suppose that* x *is bounded and let* $\overline{x} \in \operatorname{crit}(f+g)$ *be such that* $\lim_{t \longrightarrow +\infty} x(t) = \overline{x}$ *and* H *fulfills the Łojasiewicz property at* $(\overline{x}, \overline{x}, 0) \in \operatorname{crit} H$ *with Łojasiewicz exponent* θ.
Then, there exist $a_1, a_2, a_3, a_4 > 0$ *and* $t_0 > 0$ *such that for every* $t \in [t_0, +\infty)$ *the following statements are true:*

Convergence rates

(a) if $\theta \in (0, \frac{1}{2})$, then x converges in finite time;

(b) if $\theta = \frac{1}{2}$, then $\|x(t) - \overline{x}\| \leq a_1 e^{-a_2 t}$ and $\|\dot{x}(t)\| \leq a_1 e^{-a_2 t}$;

(c) if $\theta \in (\frac{1}{2}, 1)$, then $\|x(t) - \overline{x}\| \leq (a_3 t + a_4)^{-\frac{1-\theta}{2\theta-1}}$ and $\|\dot{x}(t)\| \leq (a_3 t + a_4)^{-\frac{1-\theta}{2\theta-1}}$.

- R.I. Boț, E.R. Csetnek, S.C. László, *Approaching nonsmooth nonconvex minimization through second-order proximal-gradient dynamical systems* J. Evol. Equ., (2018)
- J. Bolte, *Continuous gradient projection method in Hilbert spaces*, Journal of Optimization Theory and Applications 119(2), 235-259, (2003)
- B. Abbas, H. Attouch, *Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator*, Optimization 64(10), 2223-2252, (2015)
- R.I. Boț, E.R. Csetnek, *A forward-backward dynamical approach to the minimization of the sum of a nonsmooth convex with a smooth nonconvex function*, ESAIM: Control, Optimisation and Calculus of Variations (2018)
- A.S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, (Russian) Differentsial'nye

Uravneniya 30(9), 1475-1486, 1994; translation in Differential Equations 30(9), 1365-1375, (1994)

H. Attouch, F. Alvarez, *The heavy ball with friction dynamical system for convex constrained minimization problems*, in: Optimization (Namur, 1998), 25-35, in: Lecture Notes in Economics and Mathematical Systems 481, Springer, Berlin

Thank you for your attention!